

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

ZAVRŠNI RAD

Zrinka Murat

Zagreb, 2015. godina.

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

ZAVRŠNI RAD

**Utjecaj parametara aktivacijskih funkcija na odziv RBF
neuronske mreže**

Mentor:

Doc. dr. sc. Danko Brezak

Student:

Zrinka Murat

Zagreb, 2015. godina.

Izjavljujem da sam ovaj rad izradila samostalno, koristeći znanja stečena tijekom studija i navedenu literaturu.

Najiskrenije zahvaljujem svojem mentoru doc. dr. sc. Danku Brezaku na stručnoj pomoći, kvalitetnim smjericama, vrhunskim savjetima i poticanju interesa za ovo područje.

Ponajviše zahvaljujem svojim najmilijima koji su svakodnevno uz mene, podržavaju me i usmjeravaju te su imali razumijevanja u najzahtjevnijim trenucima mog školavanja. Hvala vam što ste tu i vjerujete u mene!

Zrinka Murat

SADRŽAJ

SADRŽAJ	I
POPIS SLIKA	II
POPIS TABLICA	III
POPIS OZNAKA	IV
POPIS KRATICA	V
SAŽETAK	VI
SUMMARY	VII
1. UVOD	8
2. UMJETNE NEURONSKE MREŽE	2
3. NEURONSKE MREŽE S RADIJALNOM BAZNOM FUNKCIJOM	4
4. AKTIVACIJSKE FUNKCIJE	8
4.1. GAUSSOVA AKTIVACIJSKA FUNKCIJA	9
4.2. INVERZNA KVADRATNA AKTIVACIJSKA FUNKCIJA	10
4.3. TROKUTNA AKTIVACIJSKA FUNKCIJA	10
5. UČENJE I TESTIRANJE	12
5.1. UČENJE	13
5.2. TESTIRANJE	15
5.3. OCJENA USPJEŠNOSTI UČENJA I TESTIRANJA	16
6. PROBLEMI NA KOJIMA JE MREŽA TESTIRANA	18
6.1. APROKSIMACIJSKI PROBLEMI	18
6.1.1. <i>Prvi aproksimacijski problem</i>	18
6.1.2. <i>Drugi aproksimacijski problem</i>	19
6.2. KLASIFIKACIJSKI PROBLEMI	20
6.2.1. <i>Krug u kvadratu (eng. circle-in-the-square)</i>	21
6.2.2. <i>IRIS primjer</i>	22
7. EKSPERIMENTALNA ANALIZA	23
7.1. ANALIZA REZULTATA DOBIVENIH NA APROKSIMACIJSKIM PROBLEMIMA	25
7.1.1. <i>Prvi aproksimacijski problem</i>	25
7.1.2. <i>Drugi aproksimacijski problem</i>	28
7.2. ANALIZA REZULTATA DOBIVENIH NA KLASIFIKACIJSKIM PROBLEMIMA	32
7.2.1. <i>Prvi klasifikacijski problem</i>	32
7.2.2. <i>Drugi klasifikacijski problem</i>	34
8. ZAKLJUČAK	37
LITERATURA	39

POPIS SLIKA

Slika 3.1.	Struktura RBF mreže	5
Slika 4.1.	Gaussova aktivacijska funkcija	9
Slika 4.2.	Inverzna kvadratna aktivacijska funkcija.....	10
Slika 4.3.	Trokutasta funkcija	11
Slika 6.1.	Prvi aproksimacijski problem	18
Slika 6.2.	Raspored točaka x,y	19
Slika 6.3.	Drugi aproksimacijski problem.....	20
Slika 6.4.	Prvi klasifikacijski problem	21
Slika 7.1.	Prikaz željenog i najboljeg dobivenog odziva na prvom aproksimacijskom problemu	26
Slika 7.2.	Ovisnost $NRMS_{test}$ o <i>koef</i> (uz Gaussovu AF) na prvom aproksimacijskom problemu	27
Slika 7.3.	Vrijednosti σ za najbolje rezultate dobivene svakom od aktivacijskih funkcija na prvom aproksimacijskom problemu.....	28
Slika 7.4.	Prikaz željenog i najboljeg dobivenog odziva na drugom aproksimacijskom problemu	29
Slika 7.5.	Ovisnost $NRMS_{testiranja}$ o <i>koef</i> (uz Gaussovu AF) na drugom aproksimacijskom problemu	30
Slika 7.6.	Detaljniji prikaz slike 7.3.	30
Slika 7.7.	Vrijednosti σ za najbolje rezultate dobivene svakom od aktivacijskih funkcija na drugom aproksimacijskom problemu.....	31
Slika 7.8.	Ovisnost UK_{test} o <i>koef</i> (uz trokutnu AF) na prvom klasifikacijskom problemu ...	33
Slika 7.9.	Vrijednosti σ za najbolje rezultate dobivene svakom od aktivacijskih funkcija na prvom klasifikacijskom problemu	34
Slika 7.10.	UK_{test} za $\beta=0,25$ a) Gaussova AF b) Trokutna AF.....	35
Slika 7.11.	Vrijednosti σ za najbolje rezultate dobivene svakom od aktivacijskih funkcija na drugom klasifikacijskom problemu	36

POPIS TABLICA

Tablica 6.1. Struktura prvog aproksimacijskog problema.....	18
Tablica 6.2. Struktura drugog aproksimacijskog problema.....	20
Tablica 6.3. Struktura prvog klasifikacijskog problema.....	21
Tablica 6.4. Struktura drugog klasifikacijskog problema.....	22
Tablica 7.1. Odabir koeficijenta β za svaki aproksimacijski i klasifikacijski problem.....	24
Tablica 7.2. Najbolji rezultati za svaki β korišten u prvom aproksimacijskom problemu.....	25
Tablica 7.3. Najbolji rezultati za svaki parametar β korišten u drugom aproksimacijskom problemu.....	28
Tablica 7.4. Najbolji rezultati za svaki parametar β korišten u prvom klasifikacijskom problemu.....	32
Tablica 7.5. Najbolji rezultati za svaki parametar β korišten u drugom klasifikacijskom problemu.....	34

POPIS OZNAKA

β	parametar ili koeficijent grupiranja
c_j	težinski faktori
C	matrica težina izlaznog sloja RBF mreže
d_{ij}	Euklidijska udaljenost
d_n	željena vrijednost izlaza
H^+	pseudoinverzija matrice H
H^{-1}	inverzija matrice H
H_{ij}	matrica radijalnih baznih funkcija
h_{ij}	vrijednost radijalne bazne funkcije
K	broj neurona skrivenog sloja
L	broj ulaznih neurona
M	broj izlaznih neurona
MS	srednja kvadratna pogreška
$NRMS$	normalizirani korijen srednje kvadratne pogreške (mjera točnosti)
$NRMS_{test}$	normalizirani korijen srednje kvadratne pogreške testiranja
$NRMS_{uč}$	normalizirani korijen srednje kvadratne pogreške učenja
O	dobivena matrica izlaznih vrijednosti
O_n	dobivena vrijednost izlaza
O_w	w-ti izlazni neuron
RMS	korijen srednje kvadratne pogreške
σ	širina (radijalne bazne) aktivacijske funkcije
t_j	vektor centara
UK	uspješnost klasifikacije
UK_{test}	uspješnost klasifikacije testiranja
$UK_{uč}$	uspješnost klasifikacije učenja
x	vektor ulaznih vrijednosti uzoraka skupa za učenje
X	matrica ulaznih vrijednosti skupa za učenje
x_i	i-ti ulazni neuron
x_{max}	ulazni podatak maksimalne vrijednosti
x_{min}	ulazni podatak minimalne vrijednosti
x_{norm}	normirana vrijednost ulaznih podataka
y	vektor izlaznih vrijednosti uzoraka skupa za učenje
y_{max}	izlazni podatak maksimalne vrijednosti
y_{min}	izlazni podatak minimalne vrijednosti
y_{norm}	normirana vrijednost izlaznih podataka

POPIS KRATICA

<i>ANN</i>	Artificial Neural Network
<i>AF</i>	Aktivacijska funkcija
<i>MS</i>	Mean Square
<i>NRMS</i>	Normalized Mean Root Square
<i>RBF</i>	Radial Basis Function
<i>RMS</i>	Root Mean Square
<i>UK</i>	Uspješnost klasifikacije

SAŽETAK

Ovaj rad temelji se na analizi karakteristika RBF neuronske mreže zasnovane na trima različitim aktivacijskim funkcijama: Gaussovoj, inverznoj kvadratnoj i trokutnoj. Generalizacijska svojstva mreže ispitana su za sve tri spomenute funkcije, te uz varijacije u iznosima parametra grupiranja i širina aktivacijskih funkcija. Procesi učenja i testiranja provedeni su korištenjem dva aproksimacijska i dva klasifikacijska problema.

Ključne riječi: RBF neuronska mreža, Gaussova funkcija, inverzna kvadratna funkcija, trokutna funkcija, grupiranje, širina aktivacijske funkcije, generalizacija.

SUMMARY

The aim of this thesis is to analyse characteristics of RBF neural network structured with the following three activation functions: Gaussian, inverse quadratic and triangular. Generalization characteristics were tested and compared for several network configurations, which were characterized not only by the types of activation functions, but also by their varying widths and positions, as well as hidden layer neurons number. Processes of learning and testing were performed on two approximationa and two classification problems.

Key words: RBF neural network, Gaussian, inverse quadratic function, triangular function, clustering, activation function width, generalization.

1. UVOD

Umjetne neuronske mreže (eng. *Artificial Neural Network*, ANN) razvijene su po uzoru na način obrađivanja informacija u ljudskom mozgu. Sastoje se od skupa međusobno povezanih umjetnih neurona čija se funkcionalnost temelji na radu biološkog neurona. Neuronske mreže zasnovane na radijalnim baznim aktivacijskim funkcijama (eng. *Radial Basis Function*, RBF) u posljednjih su petnaestak godina često korištene u rješavanju niza regresijskih i klasifikacijskih problema. Osnovne karakteristike RBF mreža su učenje u jednom koraku i jednostavna adaptacija strukture, što rezultira bržim učenjem i postizanjem potrebnih karakteristika mreže. Ovaj tip neuronske mreže spada u skupinu unaprijednih ili statičkih (eng. *Feed-Forward*) neuronskih mreža, a sastoji se od ulaznog, skrivenog i izlaznog sloja [1].

Cilj ovog rada bio je istražiti utjecaj triju različitih aktivacijskih funkcija na uspješnost učenja RBF mreže. Zato je kao podloga izrađena programska podrška, a promatrane aktivacijske funkcije bile su Gaussova funkcija, inverzna kvadratna funkcija i trokutna funkcija. Učenje, a zatim i testiranje mreže provedeno je na dva aproksimacijska i dva klasifikacijska problema.

Na uspješan rad mreže utječe i veličina skrivenog sloja, odnosno odabir broja i položaja centara neurona skrivenog sloja. S obzirom na to da ne postoji univerzalno rješenje ovog problema, uveden je parametar grupiranja β , temeljem kojeg su dobivene različite strukture skrivenog sloja mreže, odnosno generalizacijska svojstva.

Uz navedeno, postavljene strukture RBF neuronske mreže dodatno su modificirane varijacijom širina aktivacijskih funkcija kako bi se postiglo njihovo adekvatno preklapanje

Nakon toga, izvršena je analiza dobivenih rezultata, kojom su utvrđeni najbolji rezultati testiranja mreže ovisno o njejoj strukturi i parametrima aktivacijske funkcije.

2. UMJETNE NEURONSKE MREŽE

Biološku neuronsku mrežu čine biološki neuroni koji su povezani tako da obrađuju signale, a centar upravljanja im je centralni živčani sustav. Po tom uzoru izređena je umjetna neuronska mreža sastavljena od umjetnih neurona. Glavni je cilj taj da mreža pronađe prirodu povezanosti nekog ulaznog i izlaznog skupa podataka. To se odvija kroz postupak učenja koje se pokušava simulirati na onaj način na koji se odvija u ljudskom mozgu. Ipak, postoje brojna pojednostavljenja tog modela jer ljudski mozak ima vrlo kompleksnu strukturu, sastoji se od preko 100 milijardi neurona te je najneistraženiji ljudski organ [2]. Budući da nije moguće sve funkcije mozga premodelirati u matematički model, napravljena su brojna pojednostavljenja. Uz to, neke se karakteristike umjetnih neuronskih mreža ne slažu s karakteristikama biološkog neurona. Zato se može reći da je umjetni neuron dizajniran tako da oponaša osnovne funkcije biološkog neurona, a svaka im se sličnost temelji samo na procesu učenja.

Danas postoji više vrsta neuronskih mreža, no svima je zajednički postupak učenja kojeg bi se moglo objasniti kao analiziranje problema i usvajanje povezanosti između razmatranih faktora. Ovisno o pristupu, umjetne neuronske mreže mogu se podijeliti na više načina, a neki od njih su:

- Jednoslojne i višeslojne mreže (imaju ulazni, izlazni i skriveni sloj).
- Unaprijedne ili statičke (eng. *Feedforward Neural Networks*) kod kojih signali putuju u samo jednom smjeru te povratne ili dinamičke (eng. *Feedback or Recurrent Neural Networks*) kod kojih postoji barem jedna povratna petlja.
- Supervizorne, kod kojih učitelj korigira ponašanje mreže dok se ne postigne željeni rezultat i nesupervizorne koje se same organiziraju i nemaju vanjskog učitelja (samoorganizirajuće).

Neke karakteristike koje neuronske mreže posjeduju se: modeliranje nelinearnih procesa i sustava, prigušivanje šumova u signalima, brzo procesiranje podataka, rad s velikim brojem varijabli, prilagodljivost okolini i ono najvažnije, stvaranje znanja kroz učenje.

Koriste se u problemima regresije, predikcije i klasifikacije.

Danas umjetne neuronske mreže imaju primjenu u brojnim područjima, od strojarstva i fizike pa sve do medicine i bankarstva. Najčešće se koriste za raspoznavanje uzoraka, obradu podataka, donošenje odluka, probleme optimizacije, obradu slike i govora te razne simulacije. Također, važnu ulogu imaju u rješavanju problema na području umjetne inteligencije.

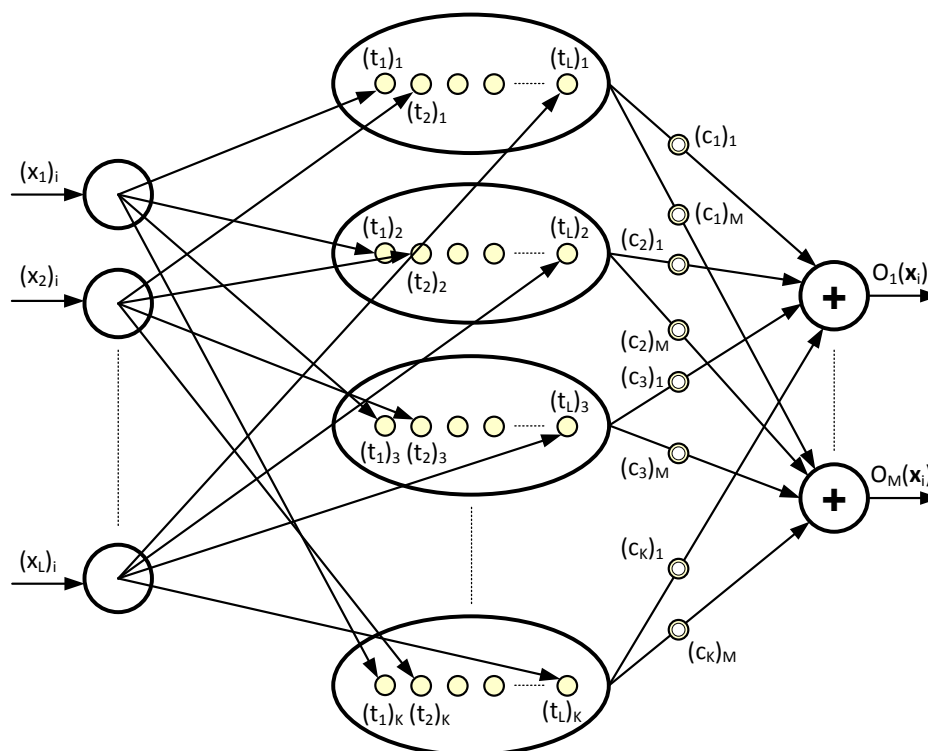
Neprestanim razvojem neuronskih mreža, razvijeni su razni algoritmi koji omogućavaju izuzetno brzo učenje mreže. Neki od njih su neiterativni postupci učenja (učenje u jednom koraku) realizirani preko inverzije ili pseudoinverzije odgovarajućih matrica sustava. Nedostatak takvog procesa učenja je izravna ovisnost dimenzija tih matrica o broju uzoraka za učenje. Ograničavanjem dimenzija matrica se ograničava i broj uzoraka za učenje, te kao rezultat proizlaze lošija generalizacijska svojstva mreže. Generalizacijska svojstva mreže označavaju sposobnost mreže da daje zadovoljavajuće rezultate na skupu ulaza različitom od onog na kojem je učila. Kako bi se tome doskočilo, koristi se kombinacija iterativnih i neiterativnih algoritama za učenje.

3. NEURONSKE MREŽE S RADIJALNOM BAZNOM FUNKCIJOM

Smisao neuronske mreže je preslikavanje zadanog ulaznog na željeni izlazni vektor. Rad s mrežom je organiziran tako da se prvo izvodi učenje, a zatim testiranje. Cilj učenja je ustanoviti kako zadani ulaz iz skupa za učenje djeluje na poznate vrijednosti njegovog izlaza. Struktura problema koji se nastoji riješiti neuronskom mrežom može imati jedan ili više ulaznih i izlaznih vektora. Kod regresijskih problema učenje se može odvijati aproksimacijom ili interpolacijom kroz zadani skup točaka za učenje. Iz tog razloga, učenje je zapravo računanje vrijednosti aproksimirane hiperravnine za zadane ulazne podatke [3]. S obzirom na nedostatke interpolacije koji uključuju problem filtriranja šuma, upitnost kvalitete učenja pri računanju odziva za uzorak koji nije element iz skupa za učenje te teško provođenje interpolacije (zbog dugotrajnog računanja i potrebne velike radne memorije računala), u ovom je radu usvojen aproksimacijski pristup.

Najbolji primjer problema s više od jedne izlazne varijable je klasifikacija, tj. zahtjev za razvrstavanjem uzoraka temeljem njihovih svojstava u različite skupine. Tada je broj neurona izlaznog sloja jednak broju klasifikacijskih skupina. U ovom je radu klasifikacija izvršena na problemu “Kvadrat u krugu” i “Iris”.

Kao što je već spomenuto, struktura mreže sastoji se od tri sloja: ulaznog, skrivenog i izlaznog. Neuroni ulaznog i skrivenog te skrivenog i izlaznog sloja potpuno su povezani, no ne postoji povezanost neurona unutar sloja u kojem se nalaze. Ulazni i izlazni sloj definirani su problemom, ulazni čini L ulaznih neurona, a izlazni M izlaznih neurona. Broj neurona skrivenog sloja (K) se određuje u fazi učenja i to od strane učitelja. Logično je da je skriveni sloj potrebno na odgovarajući način povezati s ulaznim i izlaznim slojem. Svaki neuron skrivenog sloja opisan je L -dimenzionalnim vektorom centara (t_j) i parametrima aktivacijske funkcije. Centri skrivenog sloja su povezani s pripadajućim neuronima ulaznog sloja i na taj je način ostvarena veza između ova dva sloja. Istovremeno, neuroni skrivenog sloja povezani su s neuronima izlaznog sloja preko pripadajućih težinskih faktora (c_j). Na taj se način ostvaruje preslikavanje L -dimenzionalnog ulaznog na M -dimenzionalni izlazni vektor ($\mathbb{R}^L \rightarrow \mathbb{R}^M$). Slika 3.1. prikazuje strukturu RBF mreže.



Slika 3.1. Struktura RBF mreže

Iz slike 3.1. se vidi da glavni dio RBF mreže čini upravo skriveni sloj sa svojim neuronima i težinskim faktorima. O njima ovise kvaliteta odziva i generalizacijska svojstva mreže. U slučaju interpolacije broj neurona skrivenog sloja K jednak je broju uzoraka za učenje N ($i=1, \dots, N$), a u slučaju aproksimacije broj tih neurona je manji od N [4].

Centri skrivenog sloja ovise o uzorcima za učenje i to tako što se na temelju tih uzoraka prvo računaju vrijednosti inicijalnih (početnih) centara, a onda se iste još dodatno podešavaju metodom K -srednjih vrijednosti (eng. *K-means clustering*), te se dobivaju konačne vrijednosti centara.

U izračunu inicijalnih vrijednosti centara određuje se njihov broj i početne vrijednosti, a u obzir se uzimaju svi ulazni elementi istovremeno. Računanje se provodi kroz sljedeće korake:

1. Definira se vrijednost parametra ili koeficijenta grupiranja β .
2. Kao prvi centar uzima se prvi uzorak.
3. Taj se centar uspoređuje sa svim ostalim uzorcima. Odabrana metoda uspoređivanja je izračun Euklidijske udaljenosti.

4. Dobivene udaljenosti uspoređuju se s vrijednošću β . Ako je udaljenost između promatranog uzorka i centra manja ili jednaka vrijednosti β , tada se taj uzorak svrstava u grupu s centrom s kojim je uspoređivan. Na taj se način uspoređuju svi preostali uzorci iz ulaznog skupa podataka za učenje.
5. Svi uzorci koji su na temelju ovog kriterija svrstani u istu grupu u kojoj se nalazi i centar s kojim su uspoređivani sudjeluju u izračunu centra te grupe uzoraka. Centar se računa kao srednja vrijednost razvrstanih uzoraka grupe.
6. Uzorci koji su sudjelovali u izračunu centra izuzimaju se daljnjeg izračuna početnih centara.
7. Iz reduciranog skupa uzoraka uzima se prvi uzorak, koji predstavlja početni centar druge grupe uzoraka
8. Postupak računanja centara ponavlja se od 3. koraka sve dok se svi uzorci iz ulaznog skupa podataka ne razvrstaju u svoje grupe. Na kraju se dobiva K broj centara.

Očito je da se mijenjanjem vrijednosti faktora β formira različit broj centara, a to uzrokuje promjenu strukture mreže. Stoga su u ovom radu promatrana generalizacijska svojstva mreže s obzirom na različite koeficijente β .

Nakon dobivanja početnih vrijednosti centara slijedi fino podešavanje njihovih vrijednosti. To se postiže metodom K -srednjih vrijednosti:

Računa se Euklidijska udaljenost između svakog elementa ulaznog skupa podataka i svakog prethodno izračunatog centra.

1. Za svaki centar se pronalazi minimalna Euklidijska udaljenost od svakog ulaza u mrežu, tj. uzorka za učenje.
2. Svi ulazi koji imaju minimalnu Euklidijsku udaljenost od razmatranog centra proglašavaju se članovima iste grupe.
3. Za svaku dobivenu grupu (pri čemu broj grupa iznosi K) računa se srednja vrijednost dodijeljenih uzoraka, a dobiveni rezultat proglašava se novim centrom te grupe.
4. Ukoliko su vrijednosti tako dobivenih centara jednake inicijalnim ili prethodnim vrijednostima centara postupak se prekida, te se iste proglašavaju konačnim vrijednostima centara.

U suprotnom, centar poprima novu vrijednost, a postupak se ponavlja sve dok vrijednosti novih i inicijalnih centara ne postanu jednake.

Konačne vrijednosti centara i težine izlaznog sloja oblikuju strukturu skrivenog sloja neuronske mreže te predstavljaju svojevrсно znanje neuronske mreže o učenom problemu. Dakle preostaje još pronalaženje težinskih faktora, a taj će postupak biti detaljno opisan u poglavlju 5.

4. AKTIVACIJSKE FUNKCIJE

Kao što i sam naziv mreže implicira, aktivacijska funkcija neurona skrivenog sloja mora biti odabrana iz skupine radijalnih baznih funkcija. Karakteristika ovih funkcija je njihov monoton rast ili pad s obzirom na svoje središte. Prema Micchellijevom teoremu, ako je funkcija kontinuirana u intervalu $[0, \infty)$ i pozitivna u intervalu $[0, \infty)$ te joj je prva derivacija u potpunosti monotona, tada će matrica učenja ili matrica vrijednosti radijalnih baznih funkcija imati pozitivnu determinantu. Dakle, budući da je determinanta matrice H pozitivna, biti će ostvaren uvjet njene invertibilnosti, koja je nužna za izračunavanje težinskih parametara i cjelokupno izvođenje postupka učenja. One funkcije koje zadovoljavaju taj kriterij mogu se koristiti kao radijalne bazne funkcije.

U prethodnim je poglavljima naglašena potreba povezivanja skrivenog sloja s ulaznim i izlaznim. Nakon što je objašnjena povezanost ulaznog i skrivenog sloja, slijedi prikaz veze između neurona skrivenog i izlaznog sloja:

$$O_w(x_i) = \sum_{j=1}^K h(d_{ij})c_{jw}, \quad i = 1, \dots, L, \quad w = 1, \dots, M, \quad (4.1)$$

pri čemu je O_w element w -tog izlaznog neurona koji ovisi o ulaznom vektoru x sastavljenom od svih ulaznih neurona, K broj neurona skrivenog sloja, $h(\bullet)$ aktivacijska funkcija, d_{ij} Euklidijska udaljenost između i -tog elementa ulaznog neurona i j -tog elementa vektora centra, c_{jw} težinski faktor koji povezuje j -i neuron skrivenog sloja s w -tim neuronom izlaznog sloja, L broj ulaznih, a M broj izlaznih neurona.

Iz (4.1) vidljivo je da je izlazni sloj u funkciji odziva odabrane aktivacijske funkcije.

Matricu vrijednosti radijalnih baznih funkcija H čine elementi h_{ij} , a njihova vrijednost između ostalog ovisi o odabranoj radijalnoj baznoj funkciji. Budući da je cilj ovog rada ispitati rad neuronske mreže aktivirane trima različitim funkcijama te donijeti zaključak o tome koja je pogodnija u danoj situaciji, u nastavku slijedi opis svake od njih.

4.1. Gaussova aktivacijska funkcija

Gaussova funkcija je funkcija koja prikazuje standardno odstupanje normalne razdiobe. Zvonolikog je oblika te je najčešće korištena radijalna bazna funkcija. Računa se kao

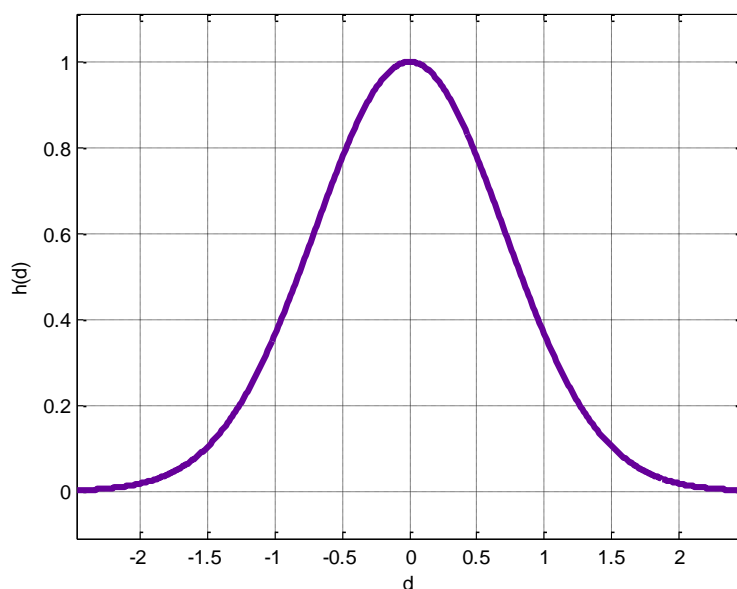
$$H(d) = e^{-\left(\frac{d}{\sigma}\right)^2}, \quad (4.2)$$

pri čemu parametar σ označava širinu radijalne bazne funkcije, a d udaljenost uzorka za učenje/testiranje od centra neurona skrivenog sloja. Vrijednost parametra σ određena je metodom P najbližih susjeda (eng. *P nearest neighbour*) [5]. Prema tome, širinu aktivacijske funkcije svakog j -tog neurona skrivenog sloja određuje izraz:

$$\sigma_j = \sqrt{d_1 d_2}, \quad (4.3)$$

gdje su d_1 , d_2 Euklidijske udaljenosti j -tog neurona od dva najbliža neurona u skrivenom sloju.

Slika 4.1. prikazuje njen izgled.



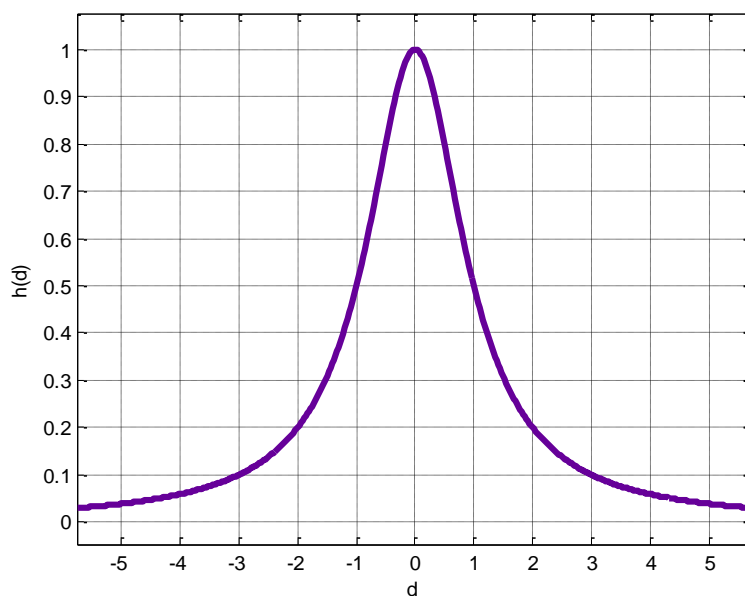
Slika 4.1. Gaussova aktivacijska funkcija

4.2. Inverzna kvadratna aktivacijska funkcija

Vrijednost inverzne aktivacijske funkcije računa se kao

$$h(d) = \frac{1}{1 + \left(\frac{d}{\sigma}\right)^2}. \quad (4.4)$$

Slika 4.2. prikazuje izgled ove funkcije.



Slika 4.2. Inverzna kvadratna aktivacijska funkcija

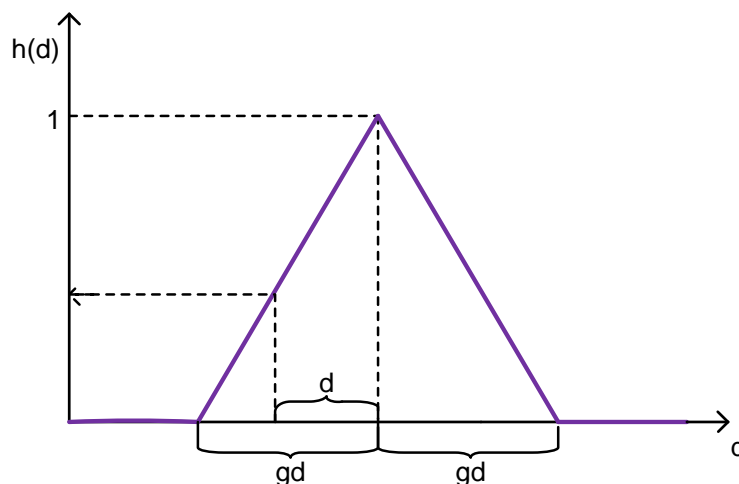
4.3. Trokutna aktivacijska funkcija

Uz uvjet da je $gd = 3\sigma$, ova aktivacijska funkcija računa se:

$$\text{za } d > gd, h(d) = 0, \quad (4.5)$$

$$\text{za } d < gd, h(d) = \left|1 - \frac{d}{gd}\right|. \quad (4.6)$$

Slika 4.3. prikazuje njen izgled.



Slika 4.3. Trokutasta funkcija

Iz prikaza ovih aktivacijskih funkcija vidi se da su sve tri funkcije simetrične te da imaju ekstremnu vrijednost za udaljenost $d=0$. Upravo to i karakterizira radijalne bazne funkcije, što znači da zadovoljavaju postavljeni uvjet.

S obzirom na to da parametar širine σ mora biti takav da postoji određeno preklapanje susjednih Gaussovih funkcija, valja biti posebno pažljiv pri njegovom odabiru. Naime, ukoliko se odabere premalena vrijednost σ , aktivacijske funkcije će biti preuske pa neće prekriti cijelo područje učenja. Direktna posljedica toga bila bi mogućnost adekvatnog određivanja izlaznih vrijednosti samo onih elemenata koji se nalaze u neposrednoj blizini centra razmatrane aktivacijske funkcije, dok bi za ostale elemente odziv mreže bio znatno lošiji. S druge strane, prevelike vrijednosti σ čine aktivacijsku funkciju preširokom, čime se gube lokalna svojstva mreže, a mogu se javiti i problemi u izračunu inverzije/pseudoinverzije matrice učenja H .

Kako bi se ispitalo mogu li se određenim modifikacijama dobiti još bolja svojstva mreže, u ovom je radu parametar σ pomnožen koeficijentom *coef* tako da sada vrijedi da je:

$$\sigma_j = coef \cdot \sqrt{d_1 d_2}. \quad (4.7)$$

Kao vrijednost ovog koeficijenta uzet je raspon brojeva $[0.1, 0.2, 0.3, \dots, 4.9, 5.0]$.

Utjecaj koeficijenta na širinu aktivacijske funkcije, kao i prikaz onih koji rezultiraju najboljim rezultatom dan je u poglavlju Eksperimentalna analiza.

5. UČENJE I TESTIRANJE

Svrha svake neuronske mreže je kvalitetno preslikavanje ulaznih podataka na izlazne podatke, a to mora biti ispunjeno i u fazi učenja i u fazi testiranja, odnosno naknadnog korištenja mreže.

Glavni cilj procedure učenja RBF mreže jest naučiti nepoznate težine kako bi se postigao kvalitetan odziv na bilo koju ulaznu veličinu, te je kao takav u potpunosti određen problemom kojim se mreža bavi. Naravno, teži se tome da se već u učenju postignu najbolja generalizacijska svojstva, odnosno da je odziv na bilo koje ulazne veličine što kvalitetniji. Kako bi se učenje moglo uspješno provesti, prije njegovog izvođenja potrebno je prikupiti odgovarajuće i pravilne uzorke za učenje. Važno je naglasiti da će učenje biti to kvalitetnije što je veći skup podataka na kojem se ono izvodi. To je samo po sebi razumljivo jer je na taj način mreži dan veći broj kombinacija ulaznih podataka, te ona može kvalitetnije naučiti prirodu njihove pozezanosti s pripadajućim izlazima. Ipak, ne treba ni pretjerivati s veličinom skupa za učenje jer bi u tom slučaju postupak učenja bio dugotrajan, a mogao bi uzrokovati i pretreniranost (eng. *overfitting*) što bi negativno utjecalo na generalizacijska svojstva mreže.

Učenje RBF mreža pripada u skupinu metoda učenja s učiteljem (eng. *supervised learning*), a učenje se odvija u jednom koraku. Ipak, na početku učenja potrebno je provesti nekoliko iteracija kako bi se odredile vrijednosti centara neurona skrivenog sloja. Taj postupak već je opisan unutar tumačenja strukture RBF neuronske mreže.

5.1. Učenje

Učenje se temelji na skupu podataka za učenje koji se sastoji od N uzoraka (N , L -dimenzionalnih vektora x), a za svaki od njih poznata je željena izlazna vrijednost vektora y .

Postupak učenja biti će prikazan na primjeru aproksimacije funkcije s više ulaza i jednim izlazom. Takav je prikaz pogodan zbog strukture podataka za aproksimaciju korištenih u procesu učenja u programskoj podršci ovog rada. Zato se uzorci skupa za učenje matematički mogu prikazati kao par (x, y) , pri čemu su

$$x = [x_1 x_2 \dots x_L], L\text{-dimenzionalni redak matrice ulaznih vrijednosti}, \quad (5.1)$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1L} \\ x_{21} & x_{22} & \dots & x_{2L} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NL} \end{bmatrix}, NxL \text{ matrica ulaznih vrijednosti}, \quad (5.2)$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}, N\text{-dimenzionalni izlazni vektor (željeni izlazni vektor)}, \quad (5.3)$$

svakom retku matrice X pridružen je jedan redak stupčanog vektora y .

Na početku faze učenja linearno se normiraju svi ulazni i izlazni podaci skupa za učenje i to na slijedeći način:

$$x_{norm_{il}} = \frac{x_{il} - x_{min_i}}{x_{max_i} - x_{min_i}}, l = 1, 2, \dots L, i = 1, 2, \dots N. \quad (5.4)$$

$$y_{norm_{il}} = \frac{y_{il} - y_{min_i}}{y_{max_i} - y_{min_i}}, \quad (5.5)$$

pri čemu su x_{norm} normirane vrijednosti ulaznih podataka, x_{min} ulazni podatak najmanje vrijednosti, x_{max} ulazni podatak najveće vrijednosti, y_{norm} normirane vrijednosti podataka izlaznog skupa, y_{min} najmanja vrijednost izlaza, y_{max} najveća vrijednost izlaza.

Nakon toga započinje učenje parametara RBF mreže. Dva poznata postupka koji se danas koriste su:

- iterativno učenje prema Poggiu i Girosiu,
- kombinirano učenje prema Moodyu i Darkenu [5].
- Poggijev i Girosijev postupak zasniva se na optimiranju kriterijskog funkcionala metodom najvećeg gradijenta, no nedostaci su mu dugotrajno učenje i rizik zaustavljanja učenja u nekom od lokalnih minimuma. I dok se u navedenoj metodi udaljenosti između uzoraka za učenje i centara neurona skrivenog sloja računaju primjenom težinske matrice, Moody i Darken se služe Euklidijskim udaljenostima, a parametar širine aktivacijske funkcije računaju zasebno za svaki neuron skrivenog sloja.

Najjednostavniji način određivanja broja i položaja centara skrivenog sloja bio bi slučajni odabir nekih od ulaznih točaka skupa za učenje, no takav odabir ne bi dao najbolja rješenja. Iz tog razloga bolje je centre računati pomoću nekog od poznatih postupaka grupiranja.

U ovom radu provedeno je kombinirano učenje koje su predložili Moody i Darken. Ono se sastoji od učenja bez učitelja nepoznatih parametara skrivenog sloja (centara i širina radijalnih baznih funkcija) i učenja s učiteljem težina izlaznog sloja.

Kako bi se definirao skriveni sloj, prvo se metodom grupiranja preko β parametra računaju inicijalni centri sigma pa se primjenjuje metoda K srednjih vrijednosti kojom su dobiveni konačni centri skrivenog sloja neuronske mreže. Ovaj je postupak već opisan u poglavlju 3.

Zatim se računa udaljenost svakog ulaznog podatka od svakog centra skrivenog sloja. Odabrani kriterij računanja udaljenosti je Euklidijska udaljenost:

$$d_{ij} = \|x_i - t_j\|, \quad (5.6)$$

gdje je x_i označava svaki ulazni neuron, a t_j svaki neuron skrivenog sloja.

Nakon toga slijedi određivanje širina aktivacijske funkcije neurona skrivenog sloja. Ono se izvodi metodom P – najbližih susjeda, uz dodatno podešavanje od strane učitelja, prema jed. (4.7)

S obzirom na odabranu aktivacijsku funkciju, prema jed. (4.2, 4.3, 4.4), u neuronima skrivenog sloja računaju se vrijednosti radijalne bazne funkcije $h(\|x_i - t_j\|)$ te se dobiva matrica vrijednosti neurona skrivenog sloja $H_{ij} = h(\|x_i - t_j\|)$. Matrica je oblika $N \times K$.

Na temelju izlaznih podataka skupa za učenje i matrice H računaju se težinski faktori izlaznog sloja, odnosno matrica C .

Izraz (4.1.) kojim je definirana povezanost između neurona skrivenog i izlaznog sloja može se prikazati i matrično:

$$O = H C, \quad (5.7)$$

Gdje O označava matricu izlaza ($N \times M$), H matricu vrijednosti radijalnih baznih funkcija ($N \times K$), a C matricu težinskih faktora mreže ($K \times M$). Matrica težinskih faktora može se tada izračunati množenjem izraza ($O=HC$) s inverzom kvadratne matrice H s lijeve strane

$$C = H^{-1}O, \quad (5.8)$$

odnosno pseudoinverzom pravokutne matrice H

$$C = H^+O, \quad (5.9)$$

pri čemu je H^+ Moore-Penroseova pseudoinverzija matrice H računa se kao

$$H^+ = (H^T H)^{-1} H^T, \quad (5.10)$$

uz uvjet da je $\det(H^T H) \neq 0$.

5.2. Testiranje

Centri neurona skrivenog sloja i težine izlaznog sloja naučeni su u fazi učenja. Oni predstavljaju bazu znanja neuronske mreže o promatranom (učenom) problemu, te se kao takvi jednom naučeni i definirani više ne mijenjaju. Dakle, u fazi testiranja koriste se njihove već izračunate vrijednosti. S obzirom na to, faza testiranja provodi se kroz sljedeće korake:

1. Izračun Euklidijevih udaljenosti svih ulaznih podataka skupa za testiranje od svih centara skrivenog sloja.
2. Izračun nove matrice radijalnih baznih funkcija H , ovisno o odabranoj aktivacijskoj funkciji.
3. Izračun odziva (izlaza).

4. Kod korištenja RBF mreže za regresiju, broj izlaznih neurona jednak je broju izlaznih veličina koje treba izračunati. To je već spomenuto i u prikazu postupka učenja koji je primijenjen kod aproksimacije funkcije. S druge strane, u klasifikaciji je broj izlaznih neurona jednak broju skupina u koje izlazne podatke treba razvrstati. Tako se nakon izračunavanja odziva sustava, s obzirom na izračunate vrijednosti izlaznih neurona i njihovo zadovoljavanje postavljenih uvjeta, izlazni podaci razvrstavaju u skupine. Na taj se način umjesto izlaznog vektora s jednom komponentom dobiva matrica izlaznih vrijednosti.

5.3. Ocjena uspješnosti učenja i testiranja

Budući da je cilj rada ispitati uspješnost algoritma učenja mreže, kao i naknadnog testiranja potrebno je odabrati prikladnu mjeru iskazivanja točnosti. Izračun mjere temelji se na željenom i ostvarenom odzivu (izlazu) mreže. Korištenje iste mjere nužno je kako bi bila moguća usporedivost rezultata dobivenih nekim drugim algoritmom. U ovom je radu uporaba iste mjere omogućila uspoređivanje rezultata uspješnosti učenja i testiranja dobivenih korištenjem istog skupa uzoraka, ali uz primjenu jedne od triju različitih radijalnih baznih funkcija. Također, na ovaj je način moguće odrediti i koja je kombinacija parametara (širine aktivacijske funkcije i broj neurona skrivenog sloja) najpovoljnija za određenu aktivacijsku funkciju. Odabir mjere ovisi o vrsti problema (predviđanje ili klasifikacija) te o njegovoj prirodi.

Danas se najčešće koriste sljedeće tri mjere uspješnosti (točnosti) algoritma učenja:

1. Srednja kvadratna pogreška (eng. *Mean Square error*, MS error), definirana kao:

$$MS = \frac{\sum_{n=1}^N (d_n - O_n)^2}{N}, \quad (5.11)$$

2. Korijen srednje kvadratne pogreške (eng. *Root Mean Square*, RMS error):

$$RMS = \sqrt{MS} = \sqrt{\frac{\sum_{n=1}^N (d_n - O_n)^2}{N}}, \quad (5.12)$$

3. Normalizirani korijen srednje kvadratne pogreške (eng. *Normalized Root Mean Square error*, NRMS error), koji se računa kao:

$$NRMS = \frac{RMS}{\sigma_{d_n}} = \frac{\sqrt{\frac{\sum_{n=1}^N (d_n - O_n)^2}{N}}}{\sigma_{d_n}}, \quad (5.13)$$

gdje je σ_{d_n} definirana kao:

$$\sigma_{d_n} = \sqrt{\frac{1}{N} \sum_{n=1}^N (d_n - \bar{d})^2}, \quad (5.14)$$

$$\bar{d} = \frac{1}{N} \sum_{n=1}^N d_n. \quad (5.15)$$

U ovom radu usvojena mjera točnosti je normalizirani korijen srednje kvadratne pogreške, odnosno NRMS mjera točnosti. Njena je prednost bezdimenzionalna neovisnost koja osigurava neovisnost mjere o dimenzijama učenih veličina te usporedbu izvedenih algoritama učenja s drugim algoritmima, neovisno o korištenoj programskoj podršci.

Izračun NRMS-a korišten je kao mjera uspješnosti dobivenih odziva učenja ($NRMS_{učenja}$) i testiranja ($NRMS_{test}$) u oba korištena aproksimacijska problema.

U slučaju dva klasifikacijska problema, a nakon razvrstavanja izlaznih podataka u pripadajuće klasifikacijske skupine, kao mjera uspješnosti klasifikacije izračunat je postotak ispravno klasificiranih uzoraka i to pomoću formula:

$$UK_{učenja} = \frac{\text{broj uspješno klasificiranih uzoraka iz skupa za učenje}}{\text{ukupan broj uzoraka u skupu za učenje}}, \quad (5.16)$$

$$UK_{testiranja} = \frac{\text{broj uspješno klasificiranih uzoraka iz skupa za testiranje}}{\text{ukupan broj uzoraka u skupu za testiranje}}. \quad (5.17)$$

Ukupna pogreška klasifikacije definira se kao postotak svih pogrešno klasificiranih uzoraka skupa za ispitivanje.

6. PROBLEMI NA KOJIMA JE MREŽA TESTIRANA

Mreža je testirana na dva aproksimacijska i dva klasifikacijska problema. U nastavku slijedi kratki prikaz tih problema, opis ulaza, opis izlaza te veličine uzorka za učenje i testiranje koji su korišteni kod svakog problema. Detaljnije o ovim problemima može se pronaći u [1].

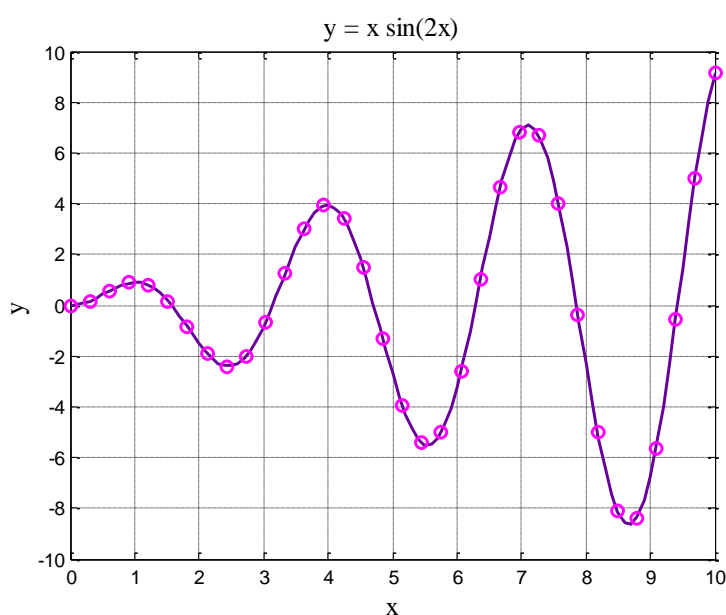
6.1. Aproksimacijski problemi

6.1.1. Prvi aproksimacijski problem

Prvi problem na kojem je mreža testirana je aproksimacija funkcije $f(x) = x \sin(x)$ u intervalu $[0, 10]$. Tablica 6.1. prikazuje strukturu podataka za učenje i testiranje, a slika 6.1. prikazuje izgled funkcije nacrtane pomoću svih podataka iz skupa za učenje.

Prvi aproksimacijski problem	
Broj ulaza:	1
Broj izlaza:	1
Broj uzoraka za učenje:	34
Broj uzoraka za testiranje:	100

Tablica 6.1. Struktura prvog aproksimacijskog problema



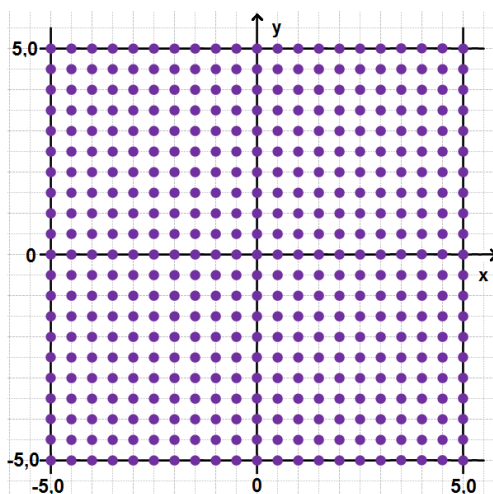
Slika 6.1. Prvi aproksimacijski problem

Iz slike 6.1. vidljivo je da su točke jednakomjerno raspoređene unutar promatranog intervala. Već je poznato da će broj centara skrivenog sloja biti manji od broja točaka iz skupa za učenje. Prema [1], pokazano je da će kod promatranog problema već deset centara ($K=10$) biti dovoljno za dobru aproksimaciju. Također, s obzirom na provedenu simulaciju potvrđeno je da je aproksimacija pogodnija za primjenu na stvarnim problemima koji su redovito opterećeni šumom.

Budući da su u ovom radu podešavani parametri β i σ , u poglavlju *Eksperimentalna analiza* biti će razmotren broj centara koji rezultira najboljom uspješnošću mreže i to s obzirom na kombinaciju promatranih parametara.

6.1.2. Drugi aproksimacijski problem

U drugom je problemu mreža učena aproksimirati funkciju $z = 2x^2\sin(y)$, pri čemu su x i y dvije nezavisne varijable. Slika 6.2. prikazuje skup podataka za učenje koji se sastoji od 441 točke ravnomjerno raspoređene unutar intervala $x, y \in [-5, 5]$.



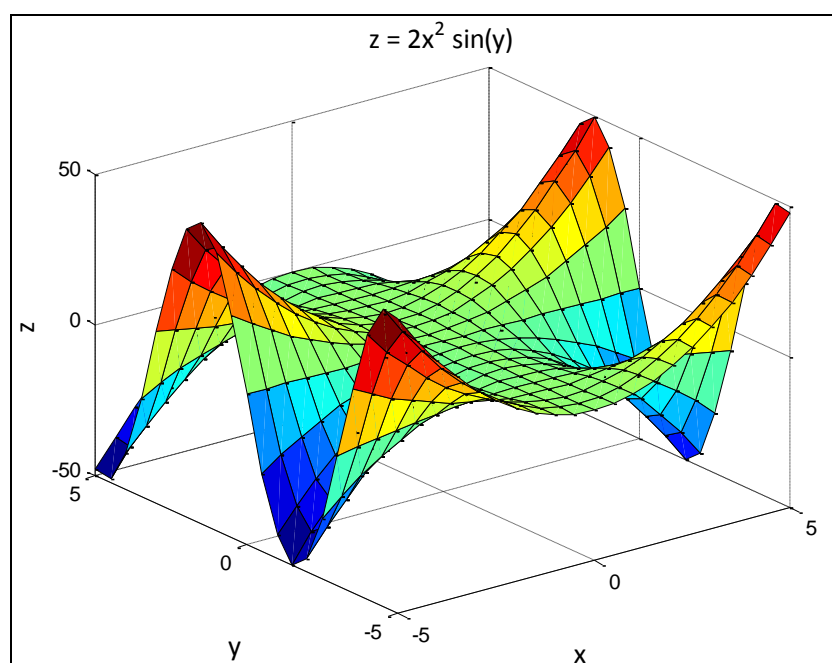
Slika 6.2. Raspored točaka x, y

Iz slike 6.2. vidljivo je da je udaljenost između točaka jednaka 0,5. S obzirom na to da je funkcija opisana dvjema varijablama, mreža ima dva ulaza i jedan izlaz. Tablica 6.2. prikazuje strukturu ovog aproksimacijskog problema.

<i>Drugi aproksimacijski problem</i>	
Broj ulaza:	2
Broj izlaza:	1
Broj uzoraka za učenje:	441
Broj uzoraka za testiranje:	900

Tablica 6.2. Struktura drugog aproksimacijskog problema

S obzirom na ulazne i izlazne podatke skupa za učenje, na slici 6.3. prikazan je izgled funkcije z .



Slika 6.3. Drugi aproksimacijski problem

6.2. Klasifikacijski problemi

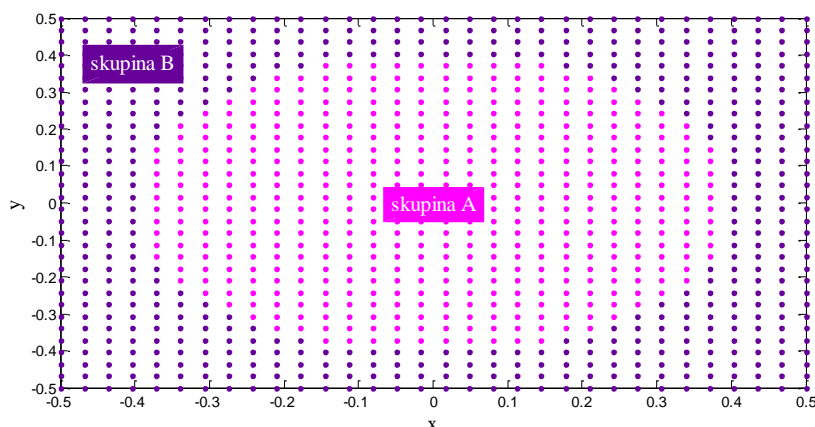
U ovom je radu funkcioniranje mreže ispitano na dva jednostavna klasifikacijska problema, “krug u kvadratu” i “IRIS”.

6.2.1. Krug u kvadratu (eng. circle-in-the-square)

Smisao ove klasifikacije jest odrediti koje se točke nalaze unutar kruga (pripadaju skupini A), a koje se nalaze izvan kruga (pripadaju skupini B), kao što prikazuje slika 6.4. Ulazni skup sastoji se od varijabli x i y koje su zapravo koordinate točaka kvadrata. Sve točke nalaze se unutar intervala $x, y \in [-0,5, 0,5]$. Izlazni skup je također dvodimenzionalan i to zato što se radi o razvrstavanju izlaznih podataka u dvije skupine. Uvjet klasifikacije je taj da je krug smješten u sredini kvadrata, sa središtem u točki $(0,0)$ te da njegova površina iznosi pola površine kvadrata. Na taj je način vjerojatnost razvrstavanja u jednu ili drugu skupinu jednaka. Za neki uzorak $x(x_1, x_2)$ uvjet se može matematički prikazati:

$$\text{ako je } r^2 = x_1^2 + x_2^2 \leq \frac{1}{2\pi} \rightarrow \text{skupina A,} \quad (6.1)$$

$$\text{ako je } r^2 = x_1^2 + x_2^2 > \frac{1}{2\pi} \rightarrow \text{skupina B.} \quad (6.2)$$



Slika 6.4. Prvi klasifikacijski problem

Tablica 6.3. prikazuje strukturu podataka na kojima je provedena klasifikacija problema “Krug u kvadratu”.

Prvi klasifikacijski problem	
Broj ulaza:	2
Broj izlaza:	2
Broj uzoraka za učenje:	1024
Broj uzoraka za testiranje:	1000

Tablica 6.3. Struktura prvog klasifikacijskog problema

6.2.2. IRIS primjer

Druga klasifikacija izvršena je pomoću Andersonovog IRIS primjera. Naime, IRIS baza podataka se sastoji od 150 uzoraka opisanih pomoću četverodimenzionalnih vektora. Svaki vektor opisuje 4 značajke cvijeta (duljinu i širinu latica cvijeta te duljinu i širinu lapova cvijeta), a cvjetovi pripadaju trima različitim skupinama cvjetova iris, setosi, versicolor i virginici. Svaka skupina opisana je s 50 uzoraka pa odatle i dolazi ukupan broj uzoraka od 150. Na taj je način ostvarena jednaka distribucija svake skupine. Opisanu strukturu prikazuje i tablica 6.4. Baza podataka korištena u ovoj klasifikaciji preuzeta je s internet stranice University of California [6].

Dakle, zadatak mreže je na temelju četverodimenzionalnog ulaznog vektora klasificirati tip cvijeta. Tablica 6.4. prikazuje strukturu problema i broj korištenih uzoraka. Vidljivo je da je i za učenje i za testiranje odabran jednak broj podataka. Podaci su uzeti tako da od 75 uzoraka (za učenje/testiranje) po 25 pripada svakoj klasi cvijeta.

Drugi klasifikacijski problem	
Broj ulaza:	4
Broj izlaza:	3
Broj uzoraka za učenje:	75
Broj uzoraka za testiranje:	75

Tablica 6.4. Struktura drugog klasifikacijskog problema

Detaljnije o provedenim simulacijama s različitim odabirom skupova za učenje i ispitivanje, različitim brojem neurona skrivenog sloja te, s obzirom na njih, različitim generalizacijskim svojstvima mreže moguće je pronaći u [1].

7. EKSPERIMENTALNA ANALIZA

Cilj ovog rada bio je usporediti karakteristike triju radijalnih baznih funkcija na dva aproksimacijska i dva klasifikacijska problema. Uz to, varijacijom broj neurona skrivenog sloja (tj. parametra β) i širina aktivacijskih funkcija (putem *koef*) tražili su se najbolji odzivi mreže.

Kako bi se taj postupak mogao provesti, napravljen je matematički model mreže u programskom paketu Matlab i to sukladno koracima opisanim u poglavlju 5. ovog rada. Mreža je izvodila učenje (i testiranje) na 4 problema, za svaki je koristila 3 aktivacijske funkcije, za svaku AF računala je 5 iteracija za različite vrijednosti parametra β i 50 iteracija za različite vrijednosti koeficijenta *koef* pomoću kojeg je mijenjana širina aktivacijske funkcije. Dakle, mreža je izvela ukupno 3000 učenja i testiranja.

Broj neurona skrivenog sloja određen je tako da se s prvom odabranom β -om dobio broj koji je za 50% manji od broja uzoraka za učenje, a sa svakom novom vrijednošću β -e broj neurona se dodatno smanjivao za pola u odnosu na prethodno stanje. No, s obzirom na različitu dinamiku svakog od četiri problema (različit broj i vrijednosti uzoraka za učenje) koji su se koristili i činjenicu da se broj neurona određuje variranjem β (a ne suprotno), nije bilo moguće odabrati 5 istih parametara β koji bi vrijedili za svaki problem. Iz tog razloga, koeficijenti β određeni su zasebno za svaki problem, no težeći tome da se barem približno postigne prvotna zamisao.

Tablica 7.1. prikazuje odabrane vrijednosti koeficijenata β za svaki problem, broj centara skrivenog sloja koji je s njima postignut te odnos aktualnog broja centara i broja uzoraka za učenje.

Aproksimacija 1		Aproksimacija 2		Klasifikacija 1		Klasifikacija 2	
β	Broj centara	β	Broj centara	β	Broj centara	β	Broj centara
0,00	34 (100%)	0,00	441 (100%)	0,00	1024 (100%)	0,00	75 (100%)
0,05	17 (50%)	0,05	225 (51%)	0,04	468 (48%)	0,09	58 (77%)
0,07	12 (35%)	0,08	121 (27%)	0,05	251 (25%)	0,13	37 (49%)
0,10	9 (26%)	0,10	85 (19%)	0,07	181 (18%)	0,25	16 (21%)
0,20	5 (15%)	0,15	42 (9%)	0,10	91 (9%)	0,50	7 (9%)

Tablica 7.1. Odabir koeficijenta β za svaki aproksimacijski i klasifikacijski problem

U nastavku je dan prikaz najboljih dobivenih rezultata za oba aproksimacijska i klasifikacijska problema i to s obzirom na odabrane parametre β . S obzirom na to da je cilj da mreža postiže što bolje rezultate u fazi testiranja, od svakog dobivenog rezultata za određeni parametar β korišten na aproksimacijskom problemu, odabran je onaj rezultat čiji je $NRMS_{testiranja}$ bio najmanji. Takav rezultat ima najmanje odstupanje dobivenog od željenog odziva. Na klasifikacijskim problemima kao najbolji rezultat dobiven uz pojedini parametar β uzet je onaj rezultat s najvećim postotokom uspješnosti klasifikacije na testiranju.

Također, biti će prikazane širine aktivacijskih funkcija kojima su postignuti najbolji rezultati na svakom od promatranih problema.

7.1. Analiza rezultata dobivenih na aproksimacijskim problemima

U nastavku slijedi analiza uspješnosti mreže pri testiranju na aproksimacijskim problemima, uz kriterij da najmanji $NRMS_{test}$ označava najbolje postignute rezultate.

7.1.1. Prvi aproksimacijski problem

U tablici 7.2. prikazani su najbolji rezultati učenja i testiranja na prvom aproksimacijskom problemu. Struktura mreže pokazuje broj neurona ulaznog, skrivenog i izlaznog sloja za svaki korišteni parametar β . Osim parametra β , za svaku AF prikazani su i oni koeficijenti *koef* koji su u kombinaciji s odgovarajućim β rezultirali najuspješnijim rezultatom testiranja.

β	Struktura mreže	Aktivacijska funkcija								
		Gaussova			Inverzna kvadratna			Trokutna		
		koef	NRMS		koef	NRMS		koef	NRMS	
			Učenje	Test		Učenje	Test		Učenje	Test
0,00	1--34--1	2,5	0,0000	0,0003	5,0	0,0001	0,0005	1,6	0,0000	0,0283
0,05	1--17--1	2,9	0,0005	0,0006	5,0	0,0048	0,0051	5,0	0,0345	0,0656
0,07	1--12--1	3,4	0,0055	0,0060	2,0	0,0224	0,0212	0,6	0,1389	0,1457
0,10	1--9--1	1,3	0,0706	0,0746	1,9	0,0999	0,1029	0,5	0,1836	0,1756
0,20	1--5--1	5,0	0,9530	0,9782	0,5	1,0037	0,9780	0,8	0,9142	0,9165

Tablica 7.2. Najbolji rezultati za svaki β korišten u prvom aproksimacijskom problemu

Iz tablice 7.2. vidi se da je najbolji rezultat $NRMS_{test} = 0,0003$. Taj je rezultat postignut s parametrom $\beta = 0,00$, odnosno u interpolacijskim uvjetima. Budući da je u ovom radu usvojen aproksimacijski pristup, u interesu je odabrati najbolji rezultat dobiven aproksimacijom, odnosno tako da je broj centara skrivenog sloja manji od veličine uzoraka na tom problemu. Iz tog je razloga kao najmanji rezultat usvojena vrijednost $NRMS_{test} = 0,0006$, a taj je rezultat postignut uz vrijednosti parametra $\beta = 0,05$ i koeficijenta *koef* = 2,9.

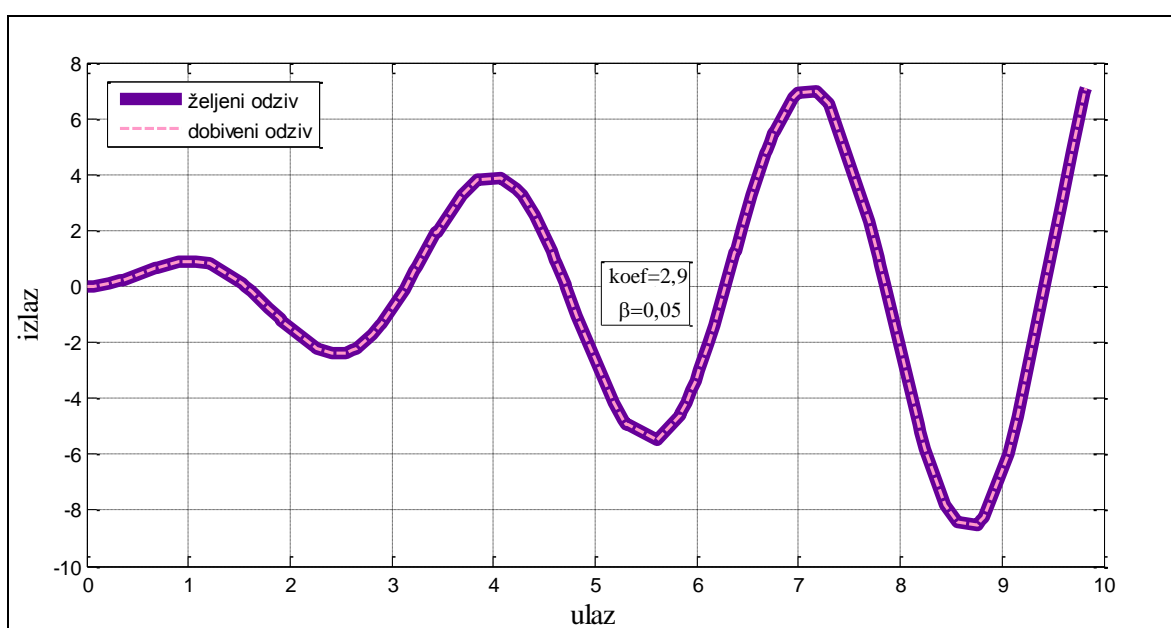
Iz prikaza rezultata vidi se da su najbolji rezultati učenja i testiranja, u ovisnosti o iznosu β , postignuti s Gaussovom, a najlošiji s trokutnom aktivacijskom funkcijom.

S porastom vrijednosti parametra β istovremeno rastu i vrijednost mjera $NRMS_{uč}$ i $NRMS_{test}$ i to tako da svaki veći β daje veću $NRMS$ vrijednost. Dakle, porastom parametra grupiranja sve

je manji broj centara skrivenog sloja, te se ostvaruju lošiji rezultati. To nije neobično iz razloga što premali broj centara nedovoljno dobro opisuje dinamiku problema pa samim time i rezultati koji se takvom mrežom mogu postići postaju lošiji.

Može se uočiti i da je većina prikazanih rezultata Gaussove i inverzne kvadratne funkcije postignuta srednjim do velikim iznosima *koef*, dok su za trokutnu funkciju u kombinaciji s većim iznosima parametra β ti *koef* nekoliko puta manji.

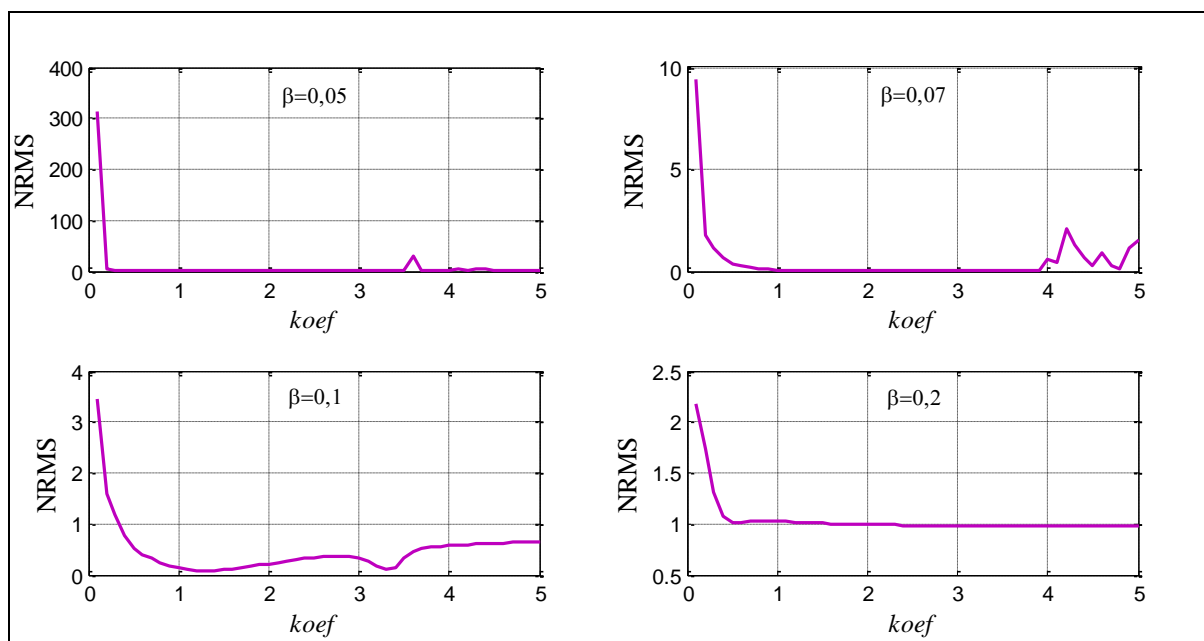
Kako bi se bolje predočio odnos *NRMS* vrijednosti i razlike između željenog i dobivenog najboljeg odziva, na slici 7.1. prikazane su dobivene i željene vrijednosti izlaza za sve ulazne podatke iz skupa za testiranje.



Slika 7.1. Prikaz željenog i najboljeg dobivenog odziva na prvom aproksimacijskom problemu

Slika 7.1. potvrđuje da je na prvom aproksimacijskom problemu Gaussovom AF uz $\beta=0,05$ i *koef*=2,9 postignut izvanredno dobar odziv mreže. Vidi se da je razlika između željnog i dobivenog odziva toliko mala da ju je teško prikazati na grafu ovih dimenzija.

Dodatno su analizirani rezultati dobiveni s Gaussovom AF i to u odnosu na sve parametre β korištene na ovom aproksimacijskom problemu. Slika 7.2. prikazuje ovisnost koeficijenta širine AF i $NRMS_{test}$ postignutog za svaki aproksimacijski parametar β , koristeći Gaussovu aktivacijsku funkciju. S obzirom na širok raspon, na slici je moguće samo uočiti trendove vrijednosti $NRMS_{test}$, dok su njihovi najmanji iznosi već prikazani u tablici 7.2.

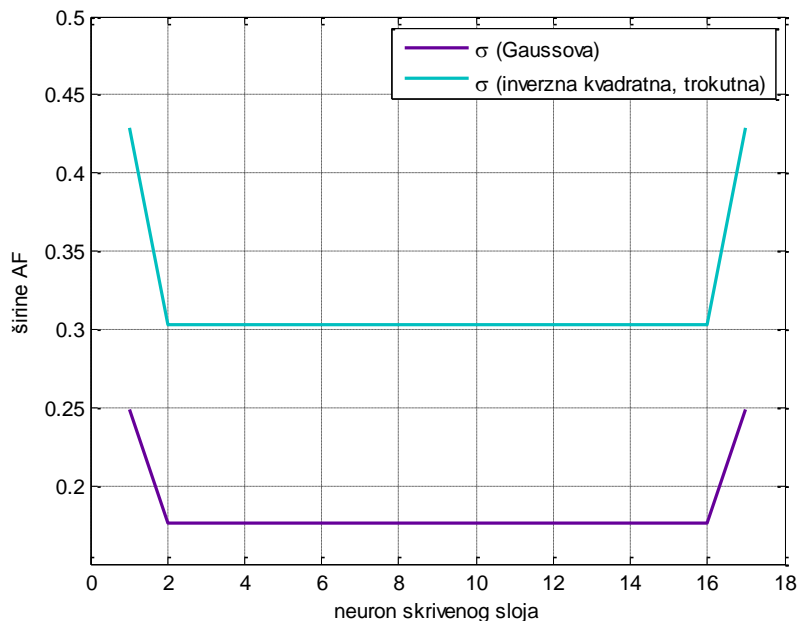


Slika 7.2. Ovisnost $NRMS_{test}$ o $koef$ (uz Gaussovu AF) na prvom aproksimacijskom problemu

Sa slike 7.2. može se vidjeti da se za $\beta=0,05$ i $koef=0,1$ postiže $NRMS_{test}=313,35$, a već sa sljedećim većim iznosom koeficijenta $koef$ počinje se naglo približavati vrijednosti 0. Vrijednost $NRMS_{test}$ ostaje takva sve do vrijednosti $koef=3,6$, te se nakon toga opet smanjuje, no više ne postiže tako niske vrijednosti kao prije. Slična situacija pojavljuje se kod vrijednosti koeficijenta $\beta=0,07$. Nakon početne vrijednosti od 9,43, $NRMS_{test}$ se naglo smanjuje te ostaje u blizini 0 sve do iznosa $koef=4,2$ te nakon toga počinju oscilacije. Traženje zakonitosti ponašanja koeficijenta $koef$ na većim iznosima β ne bi imalo smisla jer se vidi da se porastom parametra β vrijednosti $NRMS_{test}$ sve više udaljavaju od 0, te se to više ne može kompenzirati koeficijentom $koef$ širine AF.

Na slici 7.3. uspoređene su vrijednosti širina aktivacijskih funkcija za najbolje rezultate dobivene na ovom aproksimacijskom problemu. S obzirom na to da su i inverznom kvadratnom i trokutnom AF najbolji rezultati postignuti uz jednak broj neurona skrivenog sloja i uz jednake koeficijente $koef$, vrijednosti širina σ AF za svaki su centar jednake. U odnosu na njih, širina Gaussove aktivacijske funkcije duplo je manja za sve neurone skrivenog sloja.

Zajednička karakteristika ovim trima aktivacijskim funkcijama jest to da im je širina za prvi i zadnji neuron skrivenog sloja 2-3 puta veće nego za ostale neurone.



Slika 7.3. Vrijednosti σ za najbolje rezultate dobivene svakom od aktivacijskih funkcija na prvom aproksimacijskom problemu

7.1.2. Drugi aproksimacijski problem

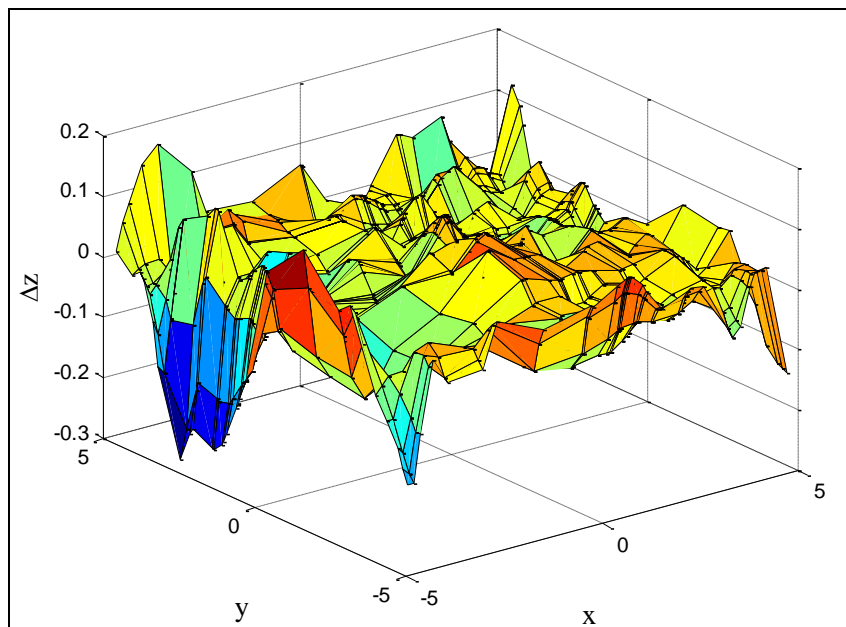
Tablica 7.3. daje prikaz najboljih rezultata ostvarenih na drugom aproksimacijskom problemu.

β	Struktura mreže	Aktivacijska funkcija								
		Gaussova			Inverzna kvadratna			Trokutna		
		koef	NRMS		koef	NRMS		koef	NRMS	
			Učenje	Test		Učenje	Test		Učenje	Test
0,00	2--441-1	2,0	0,0132	0,0227	3,7	0,0012	0,0058	2,4	0,0000	0,0675
0,05	2-225.1	2,7	0,0019	0,0030	4,7	0,0030	0,0032	4,8	0,0389	0,0511
0,08	2--121-1	2,2	0,0086	0,0112	4,2	0,0099	0,0109	4,3	0,0555	0,0658
0,10	2--85--1	2,9	0,0089	0,0084	5,0	0,0233	0,0194	5,0	0,1284	0,0932
0,15	2--42--1	3,4	0,0544	0,0470	5,0	0,0717	0,0630	5,0	0,2228	0,1739

Tablica 7.3. Najbolji rezultati za svaki parametar β korišten u drugom aproksimacijskom problemu

Iz tablice 7.3. može se uočiti da je najbolji rezultat testiranja postignut uz $\beta=0,05$ (i $koef=2,7$) te iznosi $NRMS_{test}=0,0030$. Gotovo jednako dobar rezultat postignut je i inverznom kvadratnom funkcijom (za $\beta=0,05$; $NRMS_{test}=0,0032$). Najbolji je rezultat postignut uporabom Gaussove aktivacijske funkcije.

Slika 7.4. prikazuje razliku između željenog i najboljeg dobivenog odziva na ovom problemu. Budući da se radi o problemu koji se sastoji od dva ulaza i jednog izlaza, greška aproksimacije Δz prikazana je u 3D prostoru.



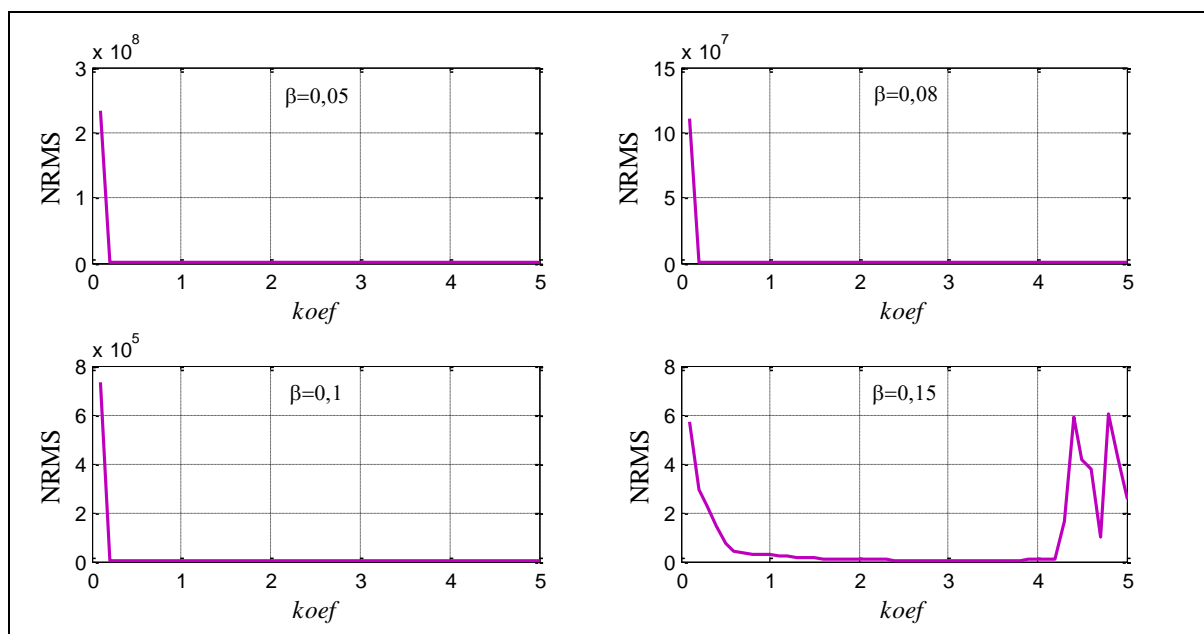
Slika 7.4. Prikaz željenog i najboljeg dobivenog odziva na drugom aproksimacijskom problemu

Zanimljivo je da za razliku od prethodnog, na ovom problemu interpolacijom nije postignut bolji rezultat nego aproksimacijom. To je dokaz tvrdnje da iako pri interpolaciji centre skrivenog sloja čine svi uzorci na kojima mreža uči, to neće nužno osigurati bolje generalizacijske karakteristike mreže.

Gaussovom su funkcijom ponovno ostvareni najbolji rezultati za svaki odabrani β , dok se trokutna funkcija pokazala najmanje uspješnom aktivacijskom funkcijom. Ono što je zanimljivo kod trokutne funkcije jest to da je kod većih vrijednosti parametra β , odnosno manjeg broja neurona skrivenog sloja, mreža ostvarila bolje rezultate na testiranju nego na učenju.

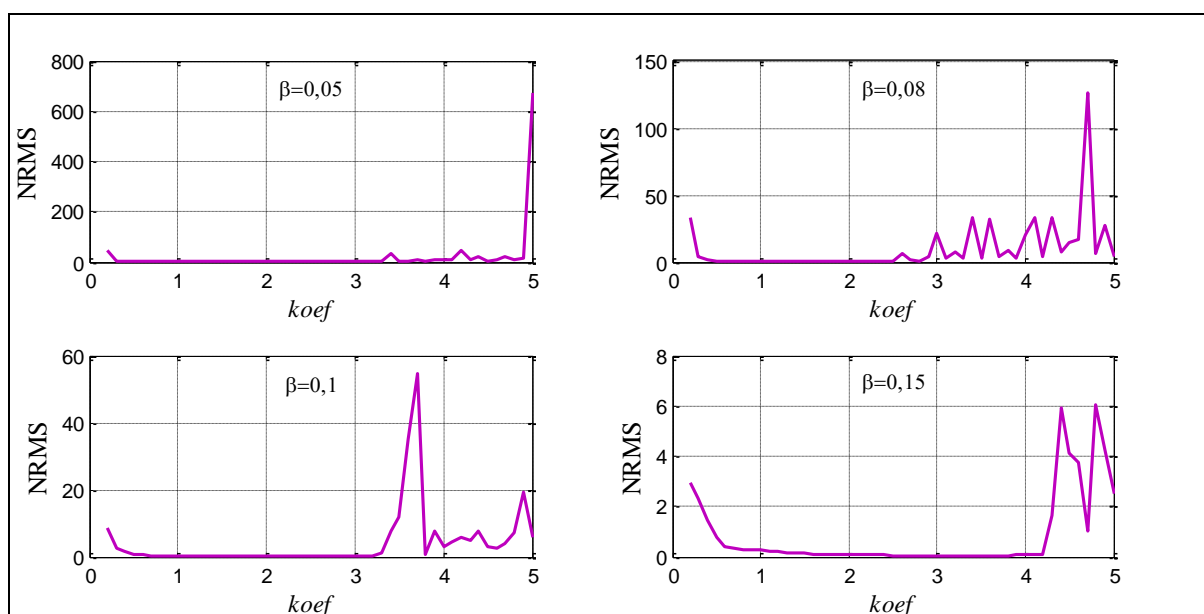
Iznosi koeficijenata *koef* kojima su postignuti najbolji rezultati, kod inverzne kvadratne i trokutne čak su duplo veći nego kod Gaussove aktivacijske funkcije.

S obzirom na to da su najbolji rezultati postignuti Gaussovom AF, slika 7.5. prikazuje ovisnost koeficijenta širine i $NRMS_{test}$ postignutog za svaki koeficijent β , koristeći Gaussovu AF.



Slika 7.5. Ovisnost $NRMS_{\text{testiranja}}$ o $koef$ (uz Gaussovu AF) na drugom aproksimacijskom problemu

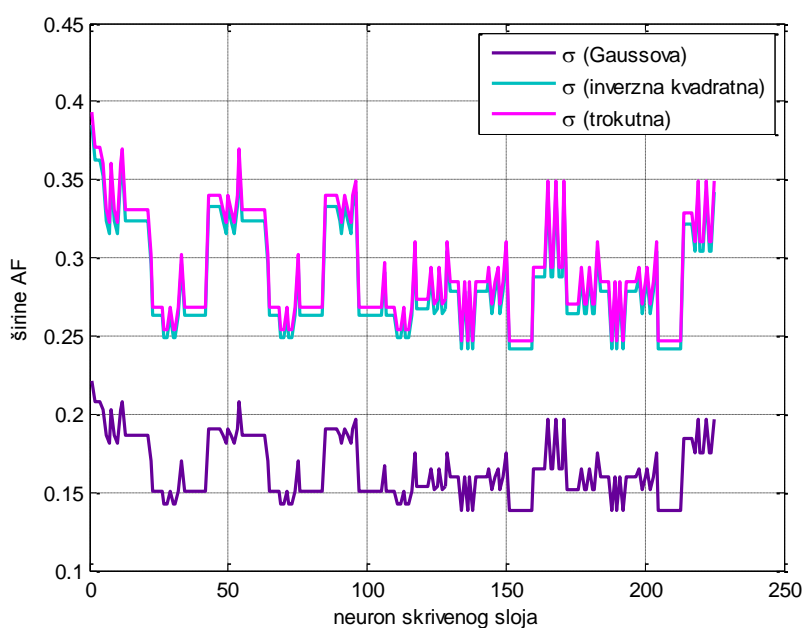
Iz slike 7.5. može se uočiti da se i ovdje pojavljuje vrijednost prvog $NRMS_{\text{test}}$ višestruko veća od preostalih vrijednosti, no za iznose $\beta=0,05$; $0,08$; $0,1$ ta je vrijednost toliko velika da je ostale iznose $NRMS_{\text{test}}$ nemoguće prikazati na finijoj skali. Kako se ne bi steknuo dojam da su sve ostale vrijednosti $NRMS_{\text{test}}$ redom jednake 0, na slici 7.6. prikazani su svi rezultati osim onog za $koef=0,1$.



Slika 7.6. Detaljniji prikaz slike 7.5.

Na slici 7.6. također se ne vide točne vrijednosti rezultata, no malo je zornije prikazano kretanje vrijednosti $NRMS_{test}$. Vidi se da se za $\beta=0,05$ vrijednosti $NRMS_{test}$ nakon naglog pada nalaze u blizini 0 i to sve do $koef=3,1$. Nakon tih vrijednosti koeficijenta $koef$ dolazi do značajnih oscilacija vrijednosti $NRMS_{test}$ od kojih je velika većina nezadovoljavajuća. Slična situacija događa se i kod većih vrijednosti parametra β pa je stoga jasno da se najbolji rezultati kod primjene Gaussove AF na drugom aproksimacijskom problemu postižu upravo najmanjim parametrom grupiranja β uz kombinaciju s ne prevelikom koeficijentom $koef$ širine aktivacijske funkcije.

Uspoređene su i vrijednosti širina aktivacijskih funkcija za najbolje rezultate testiranja dobivene na ovom problemu. Njihov prikaz nalazi se na slici 7.7 Sve tri aktivacijske funkcije postigle su najbolji rezultat uz istu vrijednost β pa je zato u sva tri slučaja jednak broj neurona skrivenog sloja. Širine trokutne i inverzne kvadratne AF vrlo se malo razlikuju, a skoro su duplo veće od širina Gaussove AF.



Slika 7.7. Vrijednosti σ za najbolje rezultate dobivene svakom od aktivacijskih funkcija na drugom aproksimacijskom problemu

I na prvom i na drugom aproksimacijskom problemu može se zaključiti da su najbolji rezultati testiranja postignuti Gaussovom aktivacijskom funkcijom uz parametar grupiranja $\beta=0,05$ te broj neurona skrivenog sloja jednak približno 50 % broja uzoraka za učenje. Također, na oba aproksimacijska problema najmanji iznos $koef=0,1$ daje vrlo loše odzive

mreže, dok samo malo veće vrijednosti postižu već zadovoljavajuće rezultate. Približavajući se vrijednosti koeficijenta $koef=3$ postižu se najbolji rezultati, a nakon toga rezultati ponovo postaju lošiji. Također, širine AF neurona skrivenog sloja koje osiguravaju zadovoljavajući odziv mreže na oba su problema puno veće za inverznu kvadratnu i trokutnu nego za Gaussovu funkciju.

7.2. Analiza rezultata dobivenih na klasifikacijskim problemima

Kao i kod aproksimacijskih problema, kriterij za najbolju strukturu mreže kod klasifikacije bio je najbolji rezultat postignut testiranjem. Razlog tome je taj što uspješno učenje mreže ne garantira jednako uspješan rad na nekim novim uzorcima.

7.2.1. Prvi klasifikacijski problem

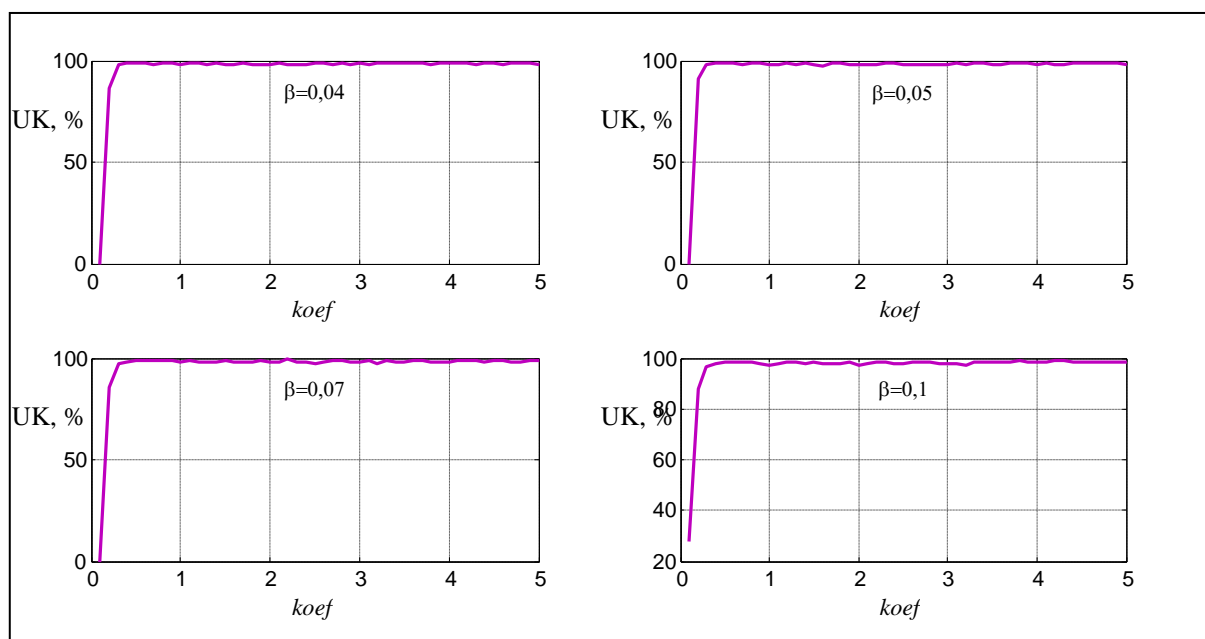
Tablica 7.4. prikazuje rezultate učenja i testiranja mreže na klasifikacijskom problemu “Krug u kvadratu”.

β	Struktura mreže	Aktivacijska funkcija								
		Gaussova			Inverzna kvadratna			Trokutna		
		koef	UK, %		koef	UK, %		koef	UK, %	
			Učenje	Test		Učenje	Test		Učenje	Test
0,00	2-1024-2	1,9	100,00	98,80	4,0	100,00	98,80	0,6	100,00	98,80
0,04	2-468-2	1,7	100,00	99,00	1,8	100,00	98,80	3,6	100,00	99,00
0,05	2-251-2	1,1	100,00	99,10	2,3	100,00	99,20	4,4	99,71	99,10
0,07	2-181-2	2,0	100,00	99,20	2,7	100,00	99,10	2,2	99,71	99,30
0,10	2-91-2	3,1	100,00	99,00	2,1	99,71	99,00	3,8	99,32	98,90

Tablica 7.4. Najbolji rezultati za svaki parametar β korišten u prvom klasifikacijskom problemu

Iz tablice 7.4. vidljivo je da je Gaussovom funkcijom za svaki iznos parametra β ostvarena stopostotna $UK_{uč}$. Inverzna kvadratna funkcija postiže istu uspješnost za četiri korištena parametra β , dok trokutna za samo dva. Usprkos tome, na testiranju, najveća uspješnost klasifikacije Gaussove i inverzne kvadratne funkcije iznosi 99%, dok trokutna funkcija ostvaruje čak 99,3%. Zanimljivost takvog rezultata je u tome što je ostvaren uz vrijednost parametra $\beta=0,07$, odnosno uz broj neurona skrivenog sloja koji iznosi 18% od ukupnog broja uzoraka za učenje.

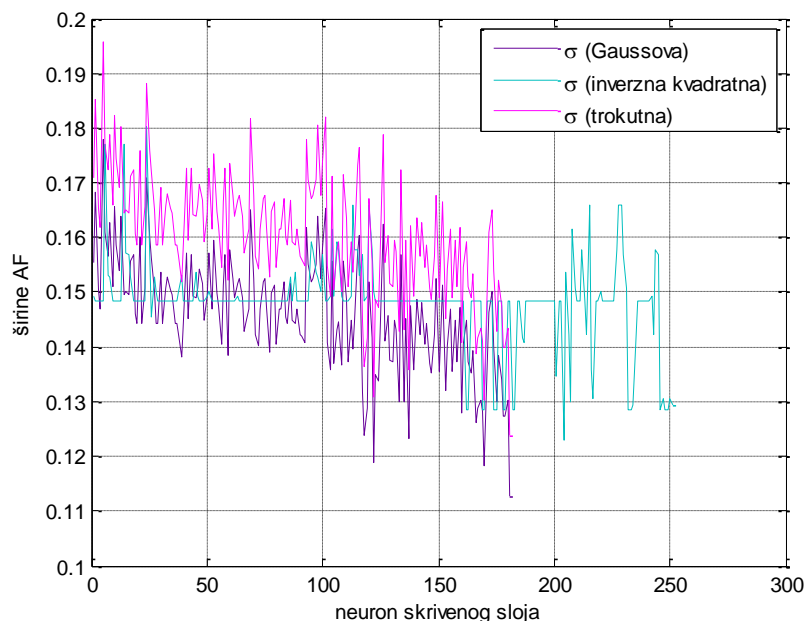
Slika 7.8. prikazuje trendove postotka uspješnosti klasifikacije za svaki β i trokutnu AF.



Slika 7.8. Ovisnost UK_{test} o $koef$ (uz trokutnu AF) na prvom klasifikacijskom problemu

Iz slike 7.8. vidi se da nakon malog početnog iznosa UK_{test} slijedi skok te se postižu zadovoljavajući rezultati. Porastom koeficijenta $koef$ vrijednosti UK_{test} i dalje ostaju velike, no pojavljuju se oscilacije.

I na ovom primjeru može se uočiti povezanost između koeficijenta $koef$ i vrijednosti σ . Zato su na slici 7.9. prikazane vrijednosti širina za svaku aktivacijsku funkciju s obzirom na njihov najbolji ostvareni odziv. Vidljivo je da za svaku aktivacijsku funkciju postoje konstantne oscilacije vrijednosti širine s obzirom na centre skrivenog sloja. Takve oscilacije očite su i u vrijednostima koeficijenata $koef$ već prikazanih u tablici 7.4. Također, jednako kao što su za trokutnu funkciju koeficijenti $koef$ najveći, a za Gaussovu najmanji, tako su i širine trokutne AF za većinu centara skrivenog sloja najveće, dok su one ostvarene Gaussovom AF najmanje.



Slika 7.9. Vrijednosti σ za najbolje rezultate dobivene svakom od aktivacijskih funkcija na prvom klasifikacijskom problemu

7.2.2. Drugi klasifikacijski problem

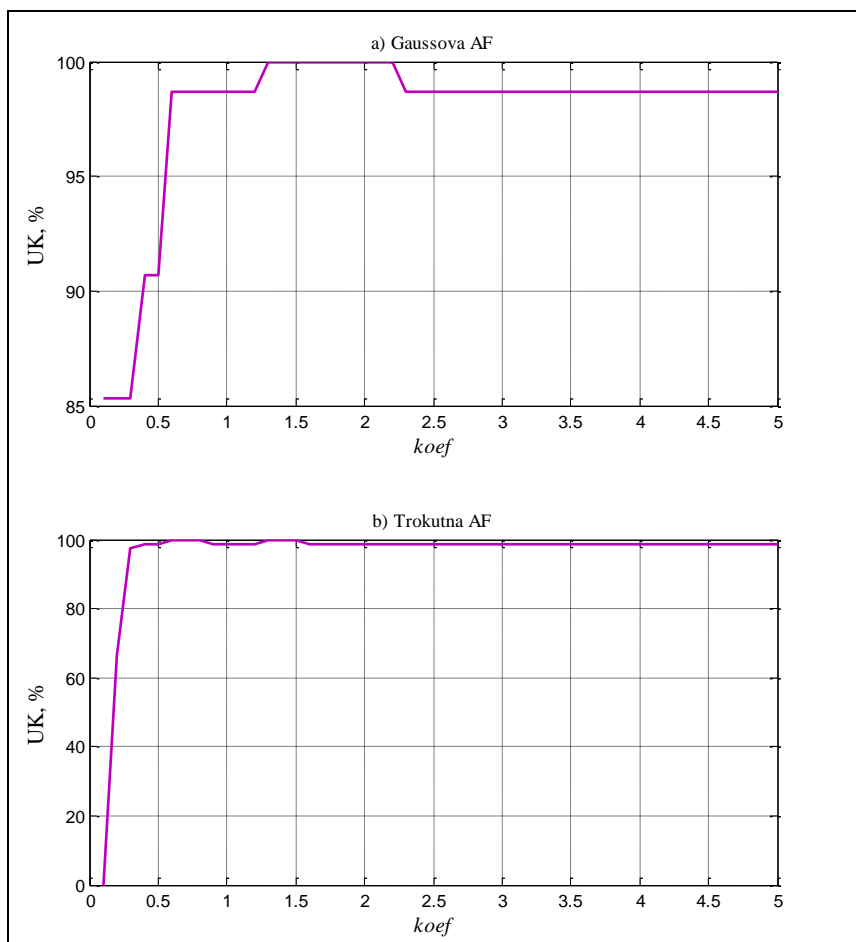
Tablica 7.5. prikazuje rezultate klasifikacije na problemu "Iris".

β	Struktura mreže	Aktivacijska funkcija								
		Gaussova			Inverzna kvadratna			Trokutna		
		koef	UK, %		koef	UK, %		koef	UK, %	
			Učenje	Test		Učenje	Test		Učenje	Test
0,00	4--75--3	0,6	100,00	97,33	0,1	100,00	97,33	2,8	100,00	98,67
0,09	4--58--3	0,1	100,00	97,33	0,3	100,00	98,67	0,9	100,00	97,33
0,13	4--37--3	1,2	98,67	98,67	0,7	97,33	98,67	1,0	97,33	100,00
0,25	4--16--3	1,3	96,00	100,00	0,6	96,00	98,67	0,6	97,33	100,00
0,50	4--7--3	2,6	97,33	97,33	4,7	96,00	97,33	0,9	94,67	93,33

Tablica 7.5. Najbolji rezultati za svaki parametar β korišten u drugom klasifikacijskom problemu

U tablici 7.5. uočava se da je kod prvih dviju najmanjih vrijednosti β ostvarena uspješnost klasifikacije učenja od 100%. Ipak, najveća uspješnost testiranja ostvarena je s nešto većim parametrima grupiranja. Naime, uspješnost klasifikacije od 100% na testiranju pojavljuje se za vrijednost $\beta=0,13$ i trokutnu aktivacijsku funkciju, ali i za vrijednost $\beta=0,25$ kod Gaussove i trokutne AF. Budući da je za parametar $\beta=0,25$ usprkos manjem broju centara skrivenog sloja postignut rezultat od 100% uspješnosti klasifikacije sa čak dvije aktivacijske funkcije,

uzima se da je to najpovoljniji parametar grupiranja primijenjen na ovom problemu. Iz tog je razloga za obje AF i $\beta=0,25$ dan prikaz ovisnosti uspješnosti klasifikacije o *koef*, Slika 7.10.

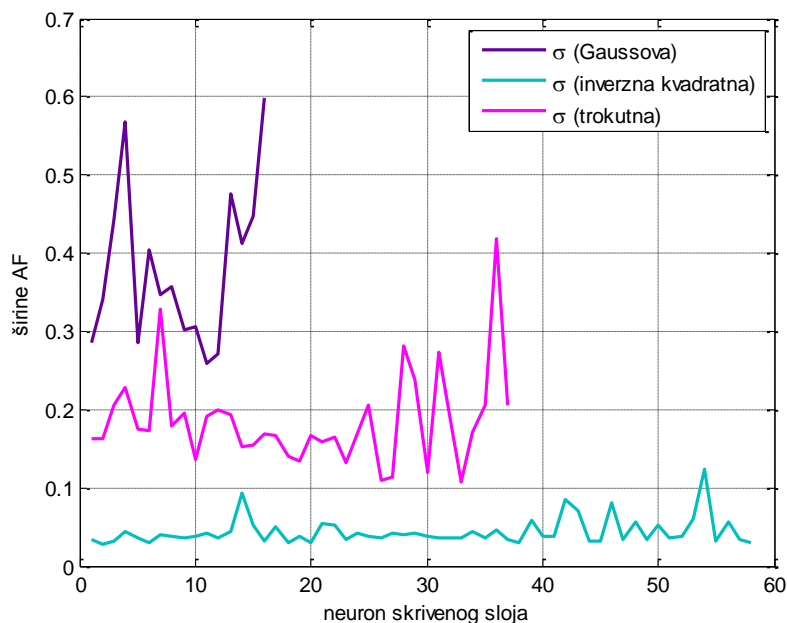


Slika 7.10. UK_{test} za $\beta=0,25$ a) Gaussova AF b) Trokutna AF

Na Slici 7.10. vidljivo je da se s Gaussovom AF postiže visoka točnost klasifikacije kod 10 različitih vrijednosti *koef*, dok se ostalim koeficijentima ostvaruje prosječan ili loš rezultat. S druge strane, trokutnom AF se $UK_{test}=100\%$ ostvaruje s manjim brojem koeficijenata (točnije njih šest), no s ostalim se vrijednostima također postiže vrlo visok postotak uspješnosti.

Uz to, budući da se $UK_{test}=100\%$ kod trokutne AF postiže s dva različita parametra β , ova se funkcija smatra najpogodnijom za primjenu na ovoj vrsti problema.

Kako bi se protumačio značaj širine AF, na slici 7.11. uspoređene su vrijednosti širina aktivacijskih funkcija za svaki najbolji rezultat dobiven na ovom klasifikacijskom problemu i to primjenom različitih aktivacijskih funkcija.



Slika 7.11. Vrijednosti σ za najbolje rezultate dobivene svakom od aktivacijskih funkcija na drugom klasifikacijskom problemu

Iz slike 7.11. vidi se da su najbolji rezultati postignuti s različitim brojem centara skrivenog sloja te da širine AF jako osciliraju od centra do centra. Jasno se vidi da su širine Gaussovih AF značajno veće nego za ostale funkcije, čak duplo veće nego u nekim centrima s trokutnom AF te 3-5 puta veće nego u nekim centrima s inverznom kvadratnom aktivacijskom funkcijom.

Konačno, najbolji rezultati klasifikacije na testiranju oba klasifikacijska problema postignuti su trokutnom funkcijom u kombinaciji s brojem centara skrivenog sloja koji iznosi približno 20% broja uzoraka za učenje.

8. ZAKLJUČAK

Nakon ispitivanja generalizacijskih svojstava mreže trima različitim aktivacijskim funkcijama, uz podešavanje parametra β za grupiranje uzoraka za učenje u centre skrivenog sloja te širina korištenih aktivacijskih funkcija dobiveno je mnoštvo različitih rezultata. Iako se radi o istom mrežnom kodu, zanimljivo je koliko spomenute varijacije utječu na promjenu dobivenih rezultata.

Ono što se odmah može zaključiti je to da nije svaka aktivacijska funkcija jednako pogodna za svaki problem kojeg mreža uči. Tako se Gaussova AF pokazala daleko najboljom za korištenje na aproksimacijskim problemima, dok je trokutna najprimjerenija za potrebe klasifikacije.

Ni najpovoljniji broj neurona skrivenog sloja ne može se univerzalno odrediti. Pokazalo se da se aproksimacijom postižu bolji rezultati s porastom broja centara. Do veće uspješnosti klasifikacije dolazi se uz nešto manji broj centara, no treba paziti da ne bude premalen.

Uz sve to, koeficijenti *koef* sadržani u širini σ aktivacijske funkcije također imaju jako velik utjecaj na generalizacijska svojstva mreže. Ukoliko je sam parametar grupiranja β bio pogodan, tada su najbolji rezultati na aproksimacijskom problemu postignuti s *koef* ne većim od 3,5. Vrlo male vrijednosti *koef* također nisu dale zadovoljavajuće rezultate. Ipak, kod klasifikacije se uz povoljno odabranu aktivacijsku funkciju zadovoljavajući rezultati postižu i s velikim vrijednostima koeficijenta *koef*.

Na razini aproksimacijskih problema zaključeno je da se s istom aktivacijskom funkcijom najbolji rezultati na oba problema postižu uz slične intervale vrijednosti njenih širina σ . S druge strane, na klasifikaciji ta zakonitost ne vrijedi.

Konačno, pokazano je da je širina trokutne i inverzne aktivacijske funkcije na većini testiranih problema veća od širine Gaussove AF. Taj trend osobito dolazi do izražaja kod aproksimacijskih problema na kojima je uočena i do 4 puta manja širine Gaussove aktivacijske funkcije.

Ovaj rad je pokazao da se svakom problemu koji se želi riješiti neuronskom mrežom treba pristupiti individualno. Naime, sama priroda problema ima vrlo velik utjecaj na rad mreže pa

je potrebno odabrati povoljne uzorke i još povoljnije parametre učenja kako bi se mreži omogućilo uspješno učenje i naknadni rad.

Na početku ovog rada povučena je analogija između umjetne neuronske mreže s biološkom neuronskom mrežom. S obzirom na kontinuirana istraživanja na području ponašanja biološkog neurona, u narednim su godinama mogući i neki novi pristupu modeliranju umjetnih neuronskih mreža, te u skladu s njima dobivanje još boljih generalizacijskih svojstava.

LITERATURA

- [1] Novaković, B., Majetić, D., Široki, M.: *Umjetne neuronske mreže*, Fakultet strojarstva i brodogradnje, Zagreb, 1998.
- [2] Wasserman, P. D.: *Neural computing: Theory and practice*, Van Nostrand Reinhold, New York, 1989.
- [3] Poggio, T., Girosi, F.: *A theory of networks for approximation and learning*, MIT, 1989.
- [4] Broomhead, D.S., Lowe, D.: *Multivariable functional interpolation and adaptive networks*, Complex Systems, Vol. 2, pp. 321-355, 1988.
- [5] Moody, J., Darken, C.: *Fast learning in networks of locally tuned processing units*, Neural Computation, Vol. 1, No. 2, pp. 281-294, 1989.
- [6] Aha, D., Murphy, P.: *UCI Repository of Machine Learning Databases and Domain Theories*, <http://mllearn.ics.uci.edu/databases/iris/>, 28.11.2014.